BF

VOLUME 6 • NUMBER 2 • APRIL 1961

CONTENTS

105 An Analysis of the Decision-Making Functions of a Simulated Air Defense Direction Center
by Anders Sweetland and William W. Haythorn

117 Instinct and History *by Nicolas Malleson*

127 A Model for Equisection Scaling *by Robert F. Fagot*

134 Models for Medical Diagnosis
by John E. Overall and Clyde M. Williams

142 A Model for the Comparison of Individual and Group Performance Scores
by Ivan D. Steiner and Nageswari Rajaratnam

CRITIQUE AND COMMENT

148 The Leviathan Technique for Large-Group Analysis
by Sydney C. Rome and Beatrice K. Rome

153 Some Remarks on the Whorfian Hypothesis
by Gerard A. Radnitzky

158 Book Reviews

COMPUTERS IN BEHAVIORAL SCIENCE

161 Preparing Psychophysiologic Analog Information for the Digital Computer *by Herbert Zimmer*

164 Automated Programming for Multinomial Distribution Problems on the IBM 650 *by James E. Farmer*

166 Computer Program Abstracts

170 Abstracts of Current Literature

173 About the Authors

BEHAVIORAL SCIENCE

A quarterly journal appearing in January, April, July, and October
Official Publication of the Mental Health Research Institute, The University of Michigan. Published at Mt. Royal & Guilford Aves., Baltimore 2, Maryland and the Mental Health Research Institute, The University of Michigan, Ann Arbor, Michigan

Behavioral Science contains articles on general theories of behavior and on empirical research specifically oriented toward such theories. An interdisciplinary approach to problems of behavior is stressed. Although the scope of the journal will include all aspects of behavior which can be subsumed under broadly general interdisciplinary theory, in the field of application special emphasis is placed on contributions relating to research in mental health and disease.

EDITORIAL BOARD: *Franz Alexander*, psychoanalysis—*Alex Bavelas*, social psychology—*Karl W. Deutsch*, political science—*David Easton*, political science—*Ralph W. Gerard*, biology (neurophysiology)—*Marion J. Levy, Jr.*, sociology—*Donald G. Marquis*, psychology—*Jacob Marschak*, economics—*James G. Miller*, psychiatry and psychology—*George P. Murdock*, anthropology—*Theodore M. Newcomb*, social psychology—*Anatol Rapoport*, mathematical biology—*Ralph W. Tyler*, education—*Raymond W. Waggoner*, psychiatry

Manuscripts should be sent in care of Dr. James G. Miller, Mental Health Research Institute, The University of Michigan, Ann Arbor, Michigan

COMPUTERS IN BEHAVIORAL SCIENCE EDITORIAL COMMITTEE: *Steven G. Vandenberg*, University of Louisville—*Bert F. Green*, Massachusetts Institute of Technology—*Bruce MacKenzie*, International Business Machines—*Charles Wrigley*, Michigan State University.

Manuscripts for the Computers in Behavioral Science department should be sent in care of Dr. Steven G. Vandenberg, Mental Health Research Institute, The University of Michigan, Ann Arbor, Michigan.

Subscription price: $6.00 a year. Price per issue: $1.75
Send subscriptions, address changes, and all business communications to *Behavioral Science*, Mental Health Research Institute, The University of Michigan, Ann Arbor, Michigan. Make all checks payable to *Behavioral Science.*

Second-class postage paid at Baltimore, Md.
Copyright, 1961, by the Mental Health Research Institute, The University of Michigan.

Shortly after the first American manual Air Defense Direction Centers began to function, it became evident that air defense was not so effective as had been hoped. The hardware was excellent; therefore, the fault must be in the way it was used. As a result, a series of studies of a simulated air defense center was begun in 1951 by the RAND Corporation. With the coming of the automatic SAGE air defense system, manual air defense has disappeared and it is now possible to declassify many findings from these studies. These facts are offered because (1) many are germane to any system involving data processing and decision making, and (2) they illustrate one way to study the decision processes of systems.

AN ANALYSIS OF THE DECISION-MAKING FUNCTIONS OF A SIMULATED AIR DEFENSE DIRECTION CENTER

by Anders Sweetland and William W. Haythorn

The RAND Corporation, Santa Monica, California

DESCRIPTION OF THE SYSTEM[1]

A MANUAL Air Defense Direction Center is a fairly autonomous Combat Information Center having the responsibility of defending a sizable piece of geography against air attack. It is composed of three closely interacting functions:

1. The Surveillance Function. Outlying early warning radars scan the horizons and (like crude television cameras) register anything that moves. This information is displayed on a radar scope that is monitored by one or more humans. These humans endeavor to separate wheat from chaff and call in the significant information. This is no easy task. The definition of "significant" varies from moment to moment (the concern of this paper). Further, a radar can, and does, faithfully report all manner of "garbage": ground clutter, storm centers, ships at sea, and (at least on one occasion) interurban trolleys. The picture is further complicated by weather phenomena (inversions), and relative reflectivity of various aircraft.

[1] Credit for the design of these experiments must go to John Kennedy, then head of the Systems Research Laboratory, William Biel, Robert Chapman, and Allen Newell. Both of the present authors came to RAND months after the last of four major studies. Most of our energies were devoted to the analysis of the data yielded by the second and third experiments. This paper is the result of part of this analysis.

All and all, the job is at once so difficult and so tedious that the men are rotated every hour to keep from "going ape."

2. The Identification Function. The reports from the "tellers" (scope watchers) are phoned into the "plotters" at the Direction Center, where the track histories are plotted on a large map of the area. Most frequently this is done by grease pencil on the back of a large illuminated plexiglass status board. When a new report comes in, the plotters draw an arrowhead with the point at the new location of the track and extend the shaft of the arrow back to the previous report. They also enter other identifying information: speed, altitude, origin, etc.

This board is continuously reviewed by the Identification Group whose major activity is to match (correlate) these tracks with flight plans that have been filed by military and commercial aircraft. Once the track is identified as friendly, it is generally "scrubbed" unless there is a chance of confusing it with other tracks. Numerous techniques are used to assist in this process: plot unknowns in red and knowns in yellow, "cease-tell" until confusion is eliminated, selective plotting, etc. When several sites simultaneously report several tracks for several minutes, the status board looks like a web made by a very confused spider.

3. The Interception Function. This is

relatively separate from the first two functions. Any tracks that remain unidentified are turned over to the Interception Controller. Interception Controller scrambles an aircraft (orders it to take off) and verbally directs it via radio out to the unknown aircraft that must be identified by type and number by the pilot of the interceptor. Because scrambles are expensive and, in cases of bad weather, dangerous, every effort is made to improve the effectiveness of the Surveillance and Identification functions to avoid a scramble.

We wish to express this process in more abstract terms. The Air Defense Center, like most decision-making centers, continually samples reality attending to those elements of information germane to its charged responsibility. These elements are used to construct a model of reality, i.e., an abstraction. This abstraction is analyzed to determine meaningful ways of dealing with reality. This determination results in a response. The process is essentially first inductive and then deductive: inductive in the construction of the model, and deductive beginning with the analysis of the model.

Concurrent and intimate to this process are a number of processes related to maintaining and improving effectiveness: 1. There is a continuous updating of the model to keep pace with the changes in reality. 2. There is a continuous process of validity checking to determine if the model is an adequate abstraction of reality. 3. There is a continuous process of technique improvement: better ways of sensing, recording, abstracting, analyzing, and responding.

PURPOSES AND METHODS

With this in mind it is possible to state the intent of the experiment. The experiment was designed to vary systematically certain dimensions of the task and observe and record the effects on the processes just described. The crews were encouraged to make as many changes in procedure as they desired. The analysis described herein was focused on discovering relationships between the variable task dimensions and crew behavior.

To protect the reader we are using a somewhat unorthodox format in this discussion: each independent variable is discussed individually (as if it were a separate experiment). It is described, its effects stated, and then discussed. Most of the numbers are omitted. The reader may assume that every statement of fact has been shown to be statistically significant. Although there are some interactions, we have included them in the discussion of the main variables to avoid confusion. These are suggestively few. Indeed, we feel that, like the barking of the hound of Baskerville, there is significance in the absence. First, a brief survey of the procedures.

The crew members were selected by the Air Defense Command with the stipulation that they be of average ability. The Cowboy crew (second experiment) consisted of 39 men. Cobra (third experiment) included 40 men. The Cobra officers, in general, were older, had higher rank, and more military and air defense experience. The Cobra noncommissioned officers had a greater proportion of higher ratings, but the same amount of military and Air Defense experience as Cowboy (average, two years).

The Direction Center was comprised of surveillance, identification, and interception functions. Three early warning sites (radars) were manned by military personnel: North, West, and South. West's zone overlapped with both North and South. These overlap areas were known as "cross-tell zones."

The experimental variables discussed here are: 1. Class of tracks: penetrating, local, or outbound. 2. Traffic load: density of air traffic. 3. Crew: Cowboy vs. Cobra. 4. Period: first vs. second half of each morning and afternoon. 5. Experience with same load: first four days with last four days. 6. Load distribution: among the three early warning sites.

The experiment consisted of four sets of four days (with a day off in between). Each day had two problems (morning and afternoon). Each problem consisted of two periods of 75 minutes each with a 40-minute nexus as the air traffic gradually changed. Thus, the design unit was the 75-minute period, of which there were 64. Each period was designed with a specific value for each

TABLE 1
FACTORIAL DESIGN OF THE STIMULUS

Load	Experience	Distribution	Period	Tracks		
				Local	Penetrating	Outbound
2	1st	Even	1	2*	2	2
			2	2	2	2
		Uneven	1	2	2	2
			2	2	2	2
	2nd	Even	1	2	2	2
			2	2	2	2
		Uneven	1	2	2	2
			2	2	2	2
3	1st	Even	1	2	2	2
			2	2	2	2
		Uneven	1	2	2	2
			2	2	2	2
	2nd	Even	1	2	2	2
			2	2	2	2
		Uneven	1	2	2	2
			2	2	2	2
4	1st	Even	1	2	2	2
			2	2	2	2
		Uneven	1	2	2	2
			2	2	2	2
	2nd	Even	1	2	2	2
			2	2	2	2
		Uneven	1	2	2	2
			2	2	2	2

* Two cases in each cell.

of the task variables. These variables were orthogonal to all other variables. This is summarized in Table 1. This is a $2 \times 3 \times 2 \times 2 \times 2 \times 3$ factorial design, with two cases per cell. The statistical approach was primarily analysis of variance.

RESULTS

The major results are summarized in Tables 2 and 3. The bottom row of Table 3 gives the percentage of total variance attributable to the significant sources of variance in the column. It is the square of the quantity (one minus remainder-mean-square divided by total-mean-square):

$$E = \left(1 - \frac{V \text{ rem}}{V \text{ tot}}\right)^2$$

This term may be interpreted essentially the same way as communality is interpreted in factor analysis: as an expression of controlled variance. As shown, most of the variance is attributable to main effects and a few first-order interactions.

One further explanation is necessary: the measures. We have used four measures of track-handling behavior.

1. *Number of Stimulus Tracks Carried (C).* A track was "carried" if the crew made a single response to it, no matter what the response was: an initial plot, position report, speed or altitude estimate, transfer or acceptance of tracking responsibility to or from another station, or increase or decrease in the number of aircraft carried under a particular track number. The number of tracks carried is thus a measure of what kind—and how many—of the many tracks in the stimulus the crew thought important enough to respond to.

2. *Percentage of Stimulus Tracks Carried (C%).* If in a certain period the stimulus contained twenty penetrating tracks and the crew carried ten of these, *C%* for penetrating tracks in that period is 50 per cent. Using the *C%* value has a number of advantages, particularly in comparing the crew's responses to different classes of tracks. On the other hand, the use of this percentage obscures changes in actual numbers (the *C* value). Fifty per cent of twenty penetrating tracks is not equivalent to 50 per cent of sixty penetrating tracks in considering the activity level of the crew.

3. *Items of Task Oriented Behavior (TOB).* Each item of track-handling information was counted once in tallying the value of *TOB*. Although this method is seemingly arbitrary, these are the units of information a crew uses in constructing its picture of a flight. Counting them gives a rough measure of the amount of crew activity.

4. *Average TOB per Track Carried (TOBav).* The average number of responses (or items of task information) per track is a measure of surveillance activity level per unit track. Changes in *TOBav* yield clues about what is causing a crew to change its procedures.

DISCUSSION

Classes of tracks

This factor accounted for most of the variance. The stimulus was built of three classes of tracks: (1) penetrating tracks, those originating outside of the system and

TABLE 2

ANALYSIS OF VARIANCE OF MEASURES OF TRACK-HANDLING BEHAVIORS†
Over-all System Performance

Source of Variance	*df*	Mean Squares for Measures of Track-Handling Behavior			
		C	*C%*	*TOB*	*TOBav*
1. Class of track	2	43,777**	108,456.5**	15,706,502**	2,671.04**
2. Load	2	872.5**	7,628.5**	70,120**	249.04**
3. Crew	1	2,070**	9,373.1**	134,767**	0.13
4. Period-within-Problem	1	481**	3,396.3**	202,142**	62.16**
5. Experience (1st-2nd)	1	94**	413.3	21,459**	13.43
6. Distribution (Even-Uneven)	1	52*	191.7	N.C.	8.14
7. Class × Load	4	2,492.3**	873.7**	351,936**	.78
8. Class × Crew	2	310**	1,261.7**	31,470**	11.45
9. Class × Period	2	70**	205.6	20,712**	10.72
10. Class × Experience	2	43.5**	279.2	6,437	5.96
11. Class × Distribution	2	25.5	177.1	N.C.	.56
12. Load × Crew	2	12	267.2	6,257	2.18
13. Load × Period	2	10.5	40.6	218	1.00
14. Load × Experience	2	24.5	67.6	9,258*	2.25
15. Load × Distribution	2	18	176.8	N.C.	3.62
16. Crew × Period	1	15	45.9	1,634	0.02
17. Crew × Experience	1	26	215.3	728	0.40
18. Crew × Distribution	1	25	145.9	N.C.	0.72
19. Period × Experience	1	5	59.6	1,911	9.04
20. Period × Distribution	1	58*	580.8*	N.C.	4.14
21. Experience × Distribution	1	28	2.2	N.C.	1.56
Remainder	253	15.05	122.44	2,894[1]	4.47
Within Cell	144	12.80	142.24	2,913[2]	4.88
Total	287	372.67	996.58	119,260	24.70

† Definition of measures: *C*—Number of tracks carried; *C%*—Percentage of stimulus tracks carried; *TOB*—Items of task oriented behavior; *TOBav*—Average *TOB* per track carried.

* $P > .05$ [1]*df*—261.

** $P > .01$ [2]*df*—216.

seen coming towards the system; (2) outbound, those going away from the system; (3) local, those remaining within the system. From the view of the participants these three classes of tracks were of differing importance. Any penetrating aircraft may be hostile (from time to time both mass raids and "sneaks"—single hostiles—were included). In contrast any track leaving the system probably is not hostile. And any originating in the system obviously is not dangerous. (The chance that an aircraft taking off from a local airport is an armed Russian bomber is, seemingly, remote.)

The differences of response between these two classes (potentially dangerous, therefore important; probably innocuous, therefore unimportant) were the numbers (and proportions) carried. Important tracks were carried, unimportant tracks dropped. This is discussed more meaningfully following the discussion of load. At the present we wish to point out that this differentiation is but one of a number of orderings along an "importance" continuum made by the crews. For example, tracks penetrating from dangerous directions (directions associated with previous hostile attacks) were carried far more often than tracks from nondangerous directions.

In general, crews maintained "important" tracks and eliminated unimportant tracks. This process was called "filtering." Our conjecture is that the stimuli were ordered along a subjective continuum varying from "no response is necessary" to "immediate response is mandatory" and that the other variables helped determine this ordering.

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE SIGNIFICANCE TEST PROBABILITIES FOR OVER-ALL SYSTEM PERFORMANCE*

Source of Variance	Measure of Track-Handling Behavior			
	*C***	*C%***	*TOB***	*TOBav***
1. Class of track	.01	.01	.01	.01
2. Load	.01	.01	.01	.01
3. Crew	.01	.01	.01	—
4. Period-within-Problem	.01	.01	.01	.01
5. Experience (1st-2ad)	.01	—	.01	—
6. Distribution (Even-Uneven)	.05	—	N.C.	—
7. Class × Load	.01	.01	.01	—
8. Class × Crew	.01	.01	.01	—
9. Class × Period	.01	—	.01	—
10. Class × Experience	05	—	—	—
11. Class × Distribution	—	—	N.C.	—
12. Load × Crew	—	—	—	—
13. Load × Period	—	—	—	—
14. Load × Experience	—	—	.05	—
15. Load × Distribution	—	—	N.C.	—
16. Crew × Period	—	—	—	—
17. Crew × Experience	—	—	—	—
18. Crew × Distribution	—	—	N.C.	—
19. Period × Experience	—	—	—	—
20. Period × Distribution	.05	.05	N.C.	—
21. Experience × Distribution	—	—	N.C.	—
Percentage of total variance attributable to significant effects	92	76	95	69

* N.C. indicates the mean square was not computed because of its obvious lack of significance. The dashes indicate nonsignificant effects.

** Definition of measures: *C* - Number of tracks carried; *C%* - Percentage of stimulus tracks carried; *TOB* - Items of task oriented behavior; *TOBav* - Average *TOB* per track carried.

This conjecture is based on the fact that most of the interactions (Tables 2 and 3) were with the class of track variable.

Load

The tracks were computed, the original information being on decks of IBM cards. A "deck" consisted of 36 tracks: 14 background ("noise") tracks, 12 penetrating, and 10 outbound. One deck was used in Load 1, two in Load 2, etc. Figure 1 shows the effect of load changes. (Figure 1 includes Loads 1 and 5. These two loads are not included in the statistical computations as they are not truly orthogonal. They were used chiefly as cross validation checks in this study.)

At first glance (Fig. 1), the crews behaved like a single-pole, 5-position switch: Load 1 yielded Load 1 behavior, Load 2 yielded Load 2 behavior, etc. This does not, however, tell all of it. The proportion of penetrating tracks from dangerous directions (and therefore potentially hostile) remained essentially the same (90–95 per cent) regardless of the load increase. But there was a marked decrease in the proportion of unimportant tracks carried (interaction of class and load).

The increase in load provoked the elimination of nonessential behavior. Load increases finally caused a pruning of almost all behavior not critical to defending the area. As will be shown, the crews also (as load went up) tended to carry tracks for shorter and shorter times, and also with fewer and fewer reports.

Thus the increase in load causes a modification of the model of reality made by the crews. At the beginning (under low loads), the crews tended to include everything (penetrating, noise, outbound) in the model. As the loads increased, they dropped the nonessentials, still maintaining an accurate replication of the essentials. This one-to-one relationship between important reality and model remained. Rather than yield, they made improvements in their processing techniques (quicker decisions, fewer reports). Our feeling is that an excellent way to measure the importance of a phenomenon to a system is to determine the accuracy of the organization's model of reality with respect to the phenomenon (the greater the accuracy the greater the importance).

Crew differences

Although there were differences between the two crews, we are inclined to minimize the significance. The chief difference was in the number of tracks carried (Cobra carried more). Closer inspection of the data showed that this was chiefly due to the number of outbound tracks carried. The least difference between the two was in the handling of the penetrating tracks: both crews maintained close, accurate contact with the stimulus on these critical tracks.

One crew (Cobra) was more authoritarian, the other (Cowboy) more democratic in approach (in the Frankel-Brunswick sense).

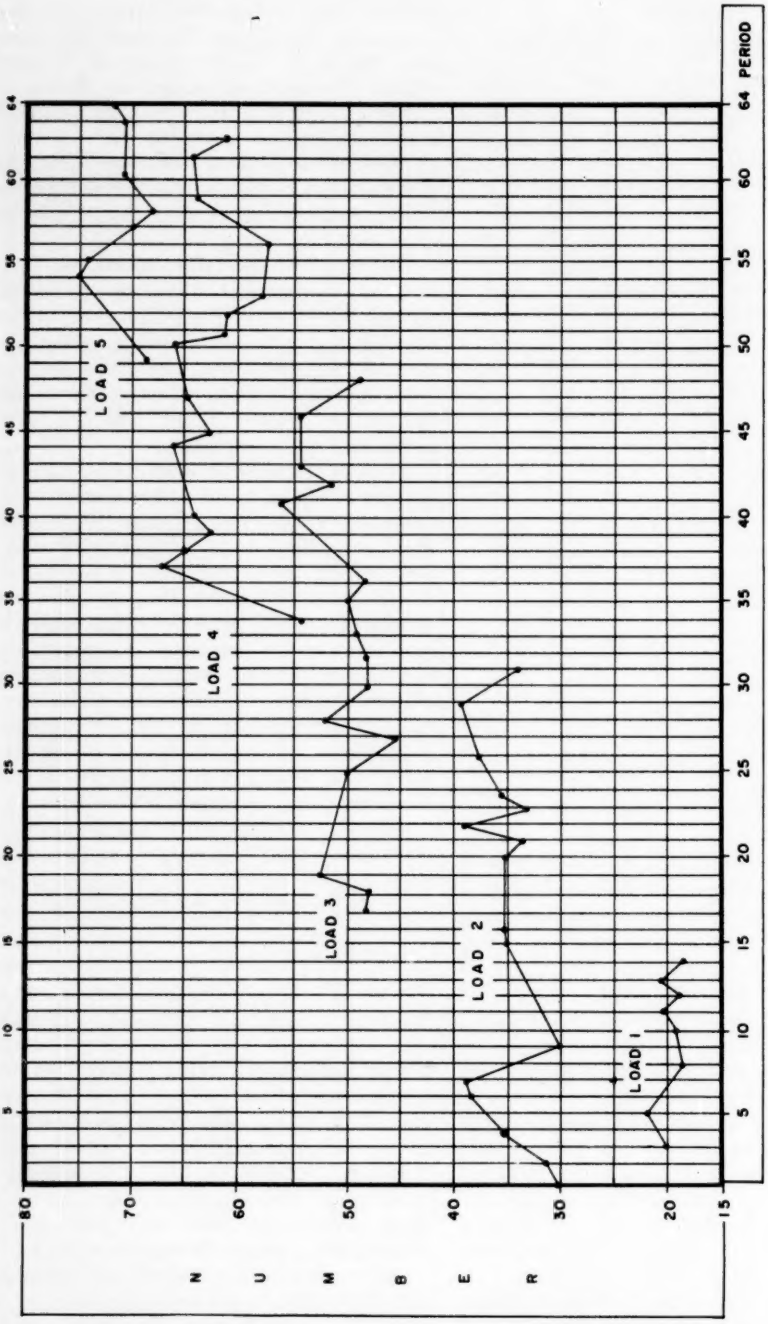

NUMBER OF PENETRATING TRACKS CARRIED
NUMBER
PERIOD
LOAD 1
LOAD 2
LOAD 3
LOAD 4
LOAD 5
80
70
60
50
40
30
20
15
5
10
15
20
25
30
35
40
45
50
55
60
64

FIGURE 1

We found some suggestions that the younger and "more democratic" Cowboy was more willing to try novel procedures which may have accounted for the fact that Cowboy drastically reduced the number of track confusions the last four days. (These increased for Cobra.) But this had little effect on over-all effectiveness.

In short, when it came to accurate handling of critical tracks, the crews performed similarly, and their behavior was essentially stimulus bound. There were some differences in the way they handled the nonessentials: the more conservative Cobra seemed less willing to prune out nonessentials and, as a result, worked harder in maintaining the same level of effectiveness. There was a slight indication that the reluctance to prune might have proved costly had the experiment gone to Load 6. It is more likely, however, that Load 6 would have provoked some hearty pruning by Cobra (i.e., as evidenced by the load by class interactions).

We cannot resist pointing out one fact: the results give no support to the champions of either the authoritarian or democratic approach. As a matter of fact, the data strongly suggest that when it comes to critical situations the question is academic; both respond to the needs of the situation, the etiquette of procedures being shelved until the crisis passes.

Period effects

One problem was run each half-day (forenoon and afternoon) and each problem consisted of two periods (the experimental unit). Comparison of the first and second periods showed greater activity (more tasks carried) during the first period. The crews were unaware of the period differences; there was a transition between them as the air traffic gradually unfolded from one situation to the other with no discontinuity.

Although a number of conjectures were offered (fatigue, boredom, motivational changes), none of the data offered substantiation. It was not until we took a more microscopic look at the data that we discovered what was happening.

At the beginning of each problem the crew was confronted with a large number of tracks that had to be sorted out, so they carried everything until it was identified (or recognized as being nonthreatening). Once they had the continuity completed, the "plot everything" behavior disappeared. Incidentally, further detailed analysis showed no evidence of a "fatigue factor." The last half hour of every day, until the close of the experiment, found the crews still actively plotting tracks at the same pace.

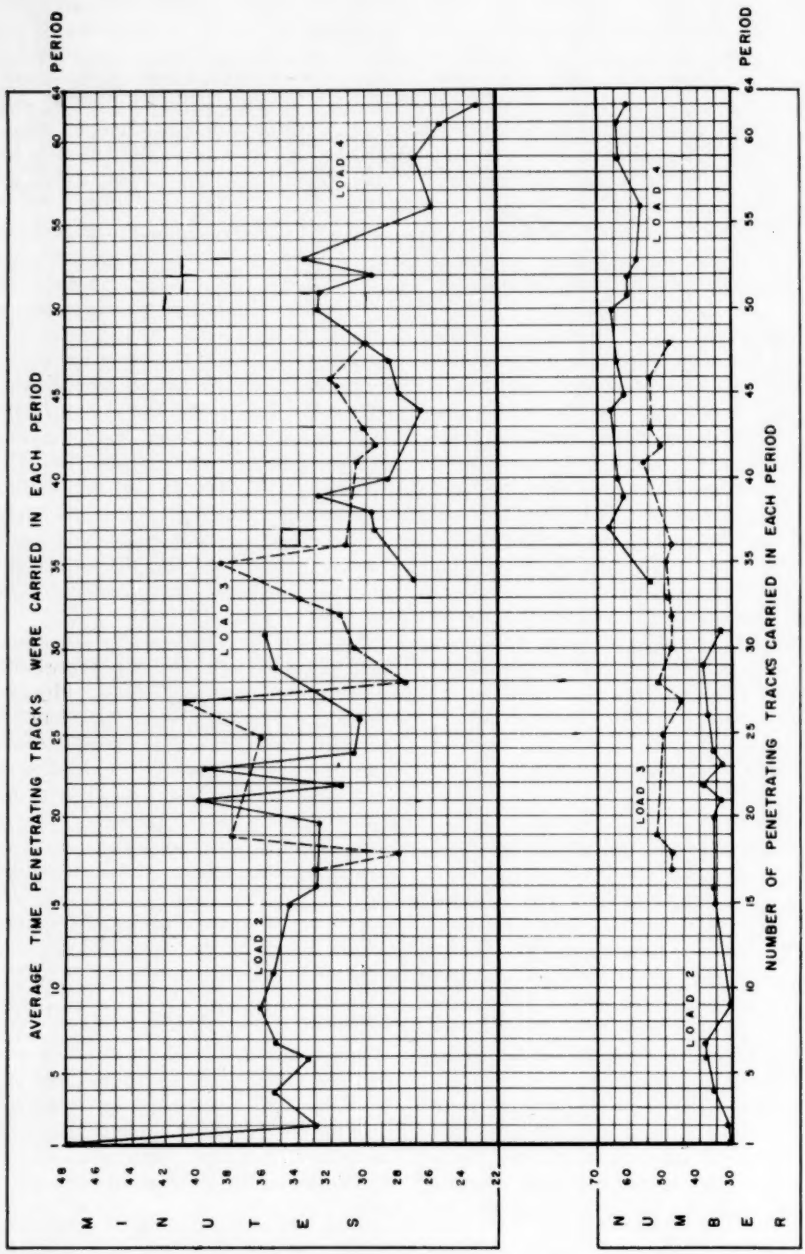
We wish to elaborate on this Period Effect. At the time, we closed our books satisfied that we had accounted for a factor peculiar to a specific simulation. However, we have noticed this same effect in a number of subsequent simulations. The effect occurs whenever a novel element is introduced into the system. At this point there is an "initial burst" activity associated with making an analysis of the new phenomenon. The characteristics of this activity are different from the processing activities: the initial burst activity is essentially diagnostic. The new phenomenon is analyzed and the results of this analysis determine the treatment (processing) to be used.

There are two reasons, in experimental work, to bear in mind the dichotomy between diagnosis and treatment. First, subjects are reluctant to proceed until they have fully analyzed the job. Accordingly, time to analyze and digest must be provided. Second, the presence of the diagnostic activity in the data contaminates the picture of the processing aspects of the operation. Consequently, it is expedient to keep these separate (if possible). An illustration of the importance of separating the two is given in the discussion of the results of the microscopic analysis.

Experience with the same load

Since Loads 2, 3, and 4 were each used in two sets, eight times as a high load in one set, then eight times as a low load in the following set, it was possible to measure the change resulting from experience (learning?) by comparing the first eight with the last eight.

These comparisons showed tiny (but statistically significant) reductions of the number of tracks carried. A few interactions



COMPARISON OF TRACKS CARRIED AND AVERAGE TIME CARRIED BY COBRA CREW

FIGURE 2

with other variables showed in a more detailed analysis, but these were erratic (occurring mostly in the data from the southern early warning station).

In general we were more impressed by the absence of evidence of learning, than the presence. We are inclined to attribute this absence to the relatively short time the crews were in the experiments as compared with their average air defense experience (two years). The microanalysis will discuss this learning factor in another dimension.

Distribution

This was included as an "even-uneven" variable. During "even" periods the load was distributed fairly evenly between the North and West early warning stations. During "uneven" periods West's load was increased while North's load was decreased. This was done to investigate "load balancing": it would be possible for North to carry all tracks in the overlap zone, when West's load was inordinate.

The data showed no definite evidence of load-balancing. Nor did any more detailed analysis show load-balancing. We have a strong suspicion that the statistical significance we obtained may be a false positive; several hundred F tests were made during the analysis. This number insures the inclusion of some "significant" results that are really random variations. Our practice has been not to attach importance to an observation unless several different measures all point toward the same interpretation. This latter did not occur with Distribution.

Further analysis

Thus far the discussion has been concerned with what was found in comparing aggregates of 75-minute periods. A number of the results suggested that a more detailed analysis was in order. Accordingly, this was done. Most of the following findings are based on data taken in 5-minute aggregates. Only one crew (Cobra) was used in this analysis (budget limitations preventing the detailed analysis of both).

The first question concerned the average time the various classes of tracks were carried under the several experimental conditions. It was found that as the experiment progressed, there was a decrease in the average time tracks were carried. The careful phrasing of the previous sentence is necessary: the design resulted in a confounding of load and experience. To determine which (if either) was predominant, the data were plotted against time (see Figure 2). The general slope (without marked separations by load) shows that, as an analysis of variance suggested, experience was the predominating factor.

A second analysis concerned the communication rate. The classes of tracks were separated and the numbers of responses were divided by the total time track-continuity was maintained. The result is a response-per-track-minute, in effect a "chatter rate." These results were almost perfect negative correlations with the previous *TOB* data; the chatter rate apparently increased with load and experience.

The close relationship between *TOB* and *C* is seen in the following. During any given five-minute period the Cobra crew carried from 2 to 17 tracks. This required the passing of a varying amount of track-handling information. To determine the relationship between *TOB* and *C*, the average amount of *TOB* per track was plotted against the number of tracks being carried (both measured in five-minute units).

The resultant curve, unusually regular for psychological data, was fitted as a log function. As can be seen in Figure 3, the data fell closely around a line fitted by the equation:

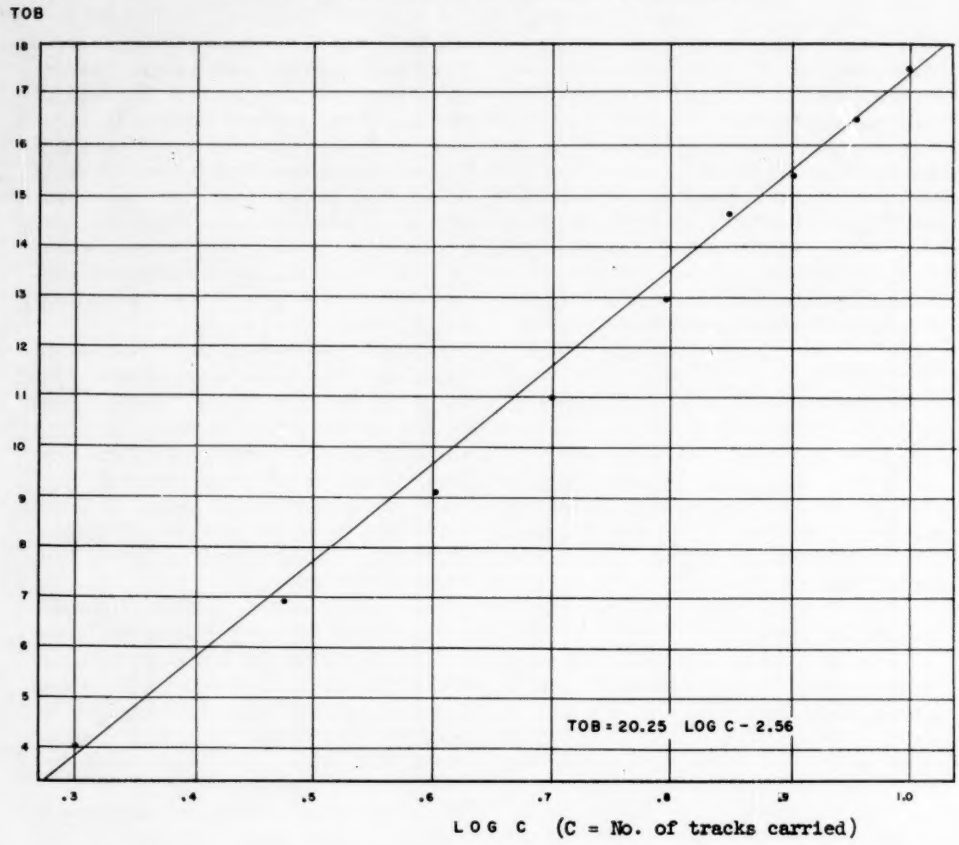
$$TOB = 20.25 (\log C) - 2.56$$

in which *TOB* is the mean number of Task-Oriented-Behaviors per five minutes and *C* is the associated number of tracks being carried.

This is interesting in view of the fact that the equation of best fit is identical to Fechner's law relating sensation to stimulus strength:

$$S = K (\log R)$$

in which S is equal to sensory magnitude, R is the stimulus measured from absolute threshold, and K is a constant. Thus the question arises as to whether it is possible to



RELATIONSHIP BETWEEN TASK ORIENTED BEHAVIOR AND NUMBER OF TRACKS CARRIED

FIGURE 3

think of *TOB* as being a measure of sensory magnitude with *C* as its stimulus.

Further study showed that when a track is initiated the crews gave: 1. an initial position report; 2. an altitude estimation; 3. a speed report; 4. a track number; and 5. a second position report (establishing direction).

Following this the track was carried by periodic position reports until it was scrubbed. Thus, a tracking activity may be divided into two fairly distinct activities: (1) initial burst and (2) carrying.

Our hypothesis that the initial burst activity caused the apparent increase in activity was tested by subtracting out the initial burst. It was then seen that the "chatter rate" remained constant throughout the experiment (i.e., that it did *not* change with load, experience, time, etc.).

The data also suggested looking at the number of simultaneous tracks carried. The results showed only a slight increase throughout the experiment. What was happening was that the crew put out essentially the same amount of effort throughout the experiment, but they were still able to handle more tracks by maintaining them under surveil-

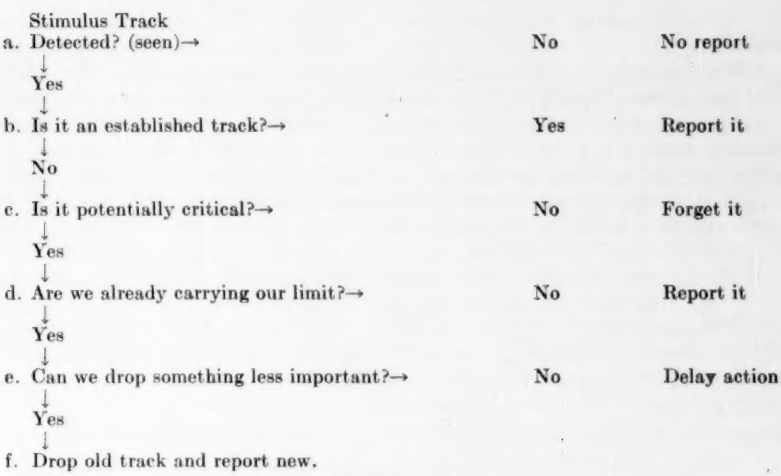


FIG. 4. Dichotomous decision model.

lance for shorter and shorter times. (The number of simultaneous tracks carried remained essentially constant: six-to-seven per plotter-teller combination. This was confirmed by analyzing the relationship between initiations and scrubs. It was found that 80 per cent of initiations occurred within two minutes following a scrub.)

With all these factors in mind it is possible to formulate a picture of what went on.

1. From the beginning the crews said, in effect: "We will carry, at any point in time, six aircraft per plotter-teller combination. Further, we will report on each aircraft once every three minutes."

2. They also said, in effect: "We will pick the six most important aircraft to maintain."

There were occasional exceptions. If fewer than six penetrating aircraft were present, they were willing to carry more than six (apparently because local and outbound tracks were easier to carry). Or, if the situation demanded, they would carry more than six criticals. The largest was 17 in a five-minute period. This happened once out of 61,000 five-minute periods.

It appears that most of these results can be explained by a series of simple, dichotomous (yes-no) decisions. This is given in Figure 4.

CONCLUSIONS

Thus, there seem to be at least three major elements in a large decision-making system:

A. Reality Modeling. When reality is too complex to handle directly, a model is made of this reality. The major characteristics of this model reflect the charged responsibility, in that the model will consist chiefly of those elements that must be controlled in order to meet the objectives of the system. The model construction is essentially an inductive process. It yields a simplified picture of reality. Its intent is of a diagnostic nature.

B. Model Analysis. The model is analyzed and the results of this analysis determine the responses of the system. These are carried out. This is essentially deductive in nature.

C. Energy Level. Prevading all this appears to be a norm that declares how hard the group or system will (or can) work to obtain the objectives. This tends to remain constant. If it happens that reality intrudes a heavier load than this normative rate permits, the least important activities are pruned, techniques are made more efficient, or better techniques are devised. In short, everything reasonable is done to keep from exceeding the system's norm. This may be due to physiological, system-structural, social, or other reasons.

One final observation: Interactions were significant by their absence. The only interactions that existed were first-order interactions, and most of these were the most obvious (i.e., interactions with class of tracks). Indeed, in doing the statistical computations, it was quickly discovered expedient to terminate computations with the first-order interactions by testing against the within groups. Almost inevitably the remaining variance was less than the within groups.

Our thesis is that, without formal training, the simplest first-order interaction is difficult for the average human being to comprehend even when the data are before him. Expecting people to show differential methods of operation in a situation as complex and amorphous as this would be expecting too much. Perhaps with months of intense experience interactions might have appeared (i.e., as they do in a thoroughly explored situation such as professional football).

It seems to us, however, that a more expedient approach (as opposed to accumulating experience) would be for systems analysts to investigate the area and, analytically, determine which areas would yield a profit if the responses were based on interactive relationships. These should be stated as decision rules: "Do X only if A is black and B is large (if A is not black or B is not large, do Y)."

(Manuscript received June 9, 1959.)

The need for authority is a constant need of man. For it is the need for principles that are both stable enough and flexible enough to give direction to the processes of living in its vicissitudes and uncertainties. Libertarians have often weakened their case by the virtual assumption that authority in every form and mode is the great enemy. In making this assumption, they play directly into the hands of those who insist upon the necessity of some external and dogmatic authority, whether ecclesiastical or political or a mixture of both. The underlying problem of recent centuries is the question of whether and how scientific method, which is the method of intelligence in experimental action, can provide the authority that earlier centuries sought in fixed dogmas. The conflict of science and religion is one phase of this conflict.

JOHN DEWEY, *Problems of Men*

The idea that man has instinctual components which shape his history and his social structures has not yet been much explored. Perhaps an historical analysis considering dominance-submission and group formation will show a recurrent pattern of events which corresponds to a pattern of subhuman primate behavior. Here historical accounts of two English revolutions and of the rise and fall of Rome provide examples of the effects upon group dynamics of "direct" and "indirect" challenges to action.

INSTINCT AND HISTORY

by Nicolas Malleson

University of London

SINCE we know for certain very little about the manifestations of instinct in man, there is still a place for the speculative hypothesis, especially if it has heuristic potentialities.

Those instincts, like hunger, thirst, and sex, which have a clear consummatory component, can be studied by the methods of physiology, psychology, and ethology. One presumes, however, that there are other instincts, in particular like those concerned with dominance-submission, with group formation, and with territorialism, which have no readily recognizable consummation. It is suggested here that these may be better studied in man in the broader fields of the social anthropologist and the historian. Their manifestation is necessarily in the interrelationships between individuals; in man these interrelationships are what constitute society. In the form and the changes of society as a whole, rather than in the behavior of individuals, instinctive elements may be clearer; on the broader canvas individual quirks may be less misleading.

It is generally agreed that in the higher mammals there is, due to learning, an increasing degree of plasticity in the manifestations of instinct. In the sort of instincts we are considering here (those that have no simple physiological consummation), this plasticity is manifest on the motor as well as the perceptual side. If the learned plasticity is assumed to be infinite in man, the study of his instincts can tell us nothing about his behavior. The conative-affective central core, to use MacDougall's phrase (1908), may be contributing some sort of nonspecific energy —call it libido or drive or what you will—but the parameters of the behavior, both in perceptual and motor aspects, will be given solely by the learning. For the purposes of this paper we will presume that this plasticity in man, though it be great, is not infinite; that patterns closer to the original adaptive form of the instinct are more readily and more usually learned than those further off. If this be the case, though one can never point a finger to the individual case and say "this is the instinctual component, this the learned," one can, given a large enough series of similar behavior patterns, make a presumption that a common instinctual component is present. There would appear to be four desiderata to be met before hazarding such a guess:

1. Comparative. To suspect a pattern of stimuli and behavior of having a definite instinctual component, it should be recognizable in species other than man. Regard should be had for taxonomic principles. Instinctual behavior of, say, fish and birds, is less valuable than that of higher mammals. The infrahuman primates should be the surest source.

2. Cross-cultural. The pattern should be recognizable in a good number of human societies, the more diverse the better. Preferably it should be recognizable in all.

3. Constellation. The regular occurrence

of a particular single stimulus or behavioral act is not by itself sufficient. The quality of instinctual behavior is given by a constellation of parts; neurophysical drive conditions, initiating stimuli, guiding stimuli, release stimuli, appetitive and taxic behavior, as well as consummatory acts and final end states. Most or all of the components must be recognizable in both the cross-cultural and comparative observations. The homosexual mounting of male baboon by male baboon at the end of a dominance-submission episode (Zuckermann, 1932) is not to be taken as homologous, in its instinctual parts, with homosexual congress among western urban human males merely because it contains a single similar act. The constellation is missing.

4. Recognizable similarities. The high degree of plasticity of man will make it unlikely that absolutely identical patterns of either perception or behavior will be found, either by cross-cultural or comparative observation. But the similarities should be recognizable.

We take it that each species has its endowment of separate instincts. However great the manifest plasticity, the conative-affective core of each presumably remains discrete, and in principle if not yet in practice, recognizable. For our purposes here we can hardly hope to recognize the influence of single instincts on a developed sequence of human behavior. We have no comprehensive list of man's endowment. Even if we had, we would have to contend with the fact that much behavior, even of lower animals, is the outcome of a synchronous interplay of more than one instinct (the baboon mounting behavior, mentioned above, which may well result from a summation of sexual instincts upon dominance instincts, the crouching submissive posture spuriously acting as a social releaser). It seems best to bypass the question altogether, and simply refer to "instinctive proclivities" to behave in such and such a manner, not concerning ourselves as to whether one or more separate instincts are involved. This will be the practice followed here.

COMPARATIVE CONSIDERATIONS

The richest source of comparative material for man is the behavior of the infrahuman primates. There have been only a few studies in anything like natural habitat: Zuckerman (1932) with baboons, Carpenter (1940) with gibbons and (1934) with howling monkeys, and Koehler (1925) with great apes. For as broad generalizations as it is proposed to make here, these seem a pitifully narrow base, but beggars can't choose.

Dominance-submission behavior, in some degree, appears common to all the infrahuman primates. Certain events seem regularly to be perceived as challenges to dominance and to call forth aggressive behavior—particular examples are the sexual approach to a possessed female, the taking of scarce food, and the trespassing into territorial preserves. (i) *In man we might expect to find instinctive proclivities to manifest dominance-submission behavior, and to respond to what is felt as a challenge by dominating aggressive behavior.*

In the infrahuman primates, the threat behavior of the dominant individual regularly calls out submissive behavior, flight or submissive posture etc., in the other. (ii) *In man we might expect to find proclivities to respond submissively to individuals felt to be dominant, thereby establishing in effect a dominance scale.*

In the infrahuman primates, the same animals that at times show dominance-submission behavior with each other also show behavior that can best be described as amicable: play, grooming, rescue responses, etc. (iii) *In man we might expect to find proclivities to engage in amicable behavior even though, at other times, dominance-submission behavior prevailed between the same individuals.*

In the infrahuman primates it is found that groups are formed, and much or all of most individual lives are lived within a group. These groups are basically family groups, but their constitutions (modal numbers of males, females, and young, etc.) vary according to the sexual and physiological characteristics of the species. In these groups both dominance-submission and amicable behavior is to be found. A fairly static domi-

nance scale, at least as far as the males are concerned, is normally present. (iv) *In man we might expect to find proclivities to form groups which encompass both dominance-submission behavior with a dominance scale, and amicable behavior.*

In the infrahuman primates the animals in a group seem to cease their own interplay, in particular their own dominance-submission disputes, when faced with a challenging event from outside the group, in particular the trespass of others on a territorial preserve, the introduction of a stranger, or the assault by an outsider on a group member. The aggressive dominating proclivities of all the group members are then usually directed outwards effectively defending group territory or members. (v) *In man we might expect to find proclivities to cease intragroup conflict and engage in group-cohesive outwardly directed and aggressive behavior in response to an event perceived as a challenge from outside the group.*

It would seem not unreasonable to look at the form and history of human societies to see whether the marks of instinctive proclivities of this kind are recognizable. At first blush, at any rate, they do not seem disconsonant with our common sense notions of men and society around us.

DOMINANCE-SUBMISSION IN HUMAN SOCIETY

In all human societies "rank" and "status" play an important part. Among men relationships are seldom equal. The inequality may be socially prescribed, policeman to barrow boy, officer to private soldier, or it may arise simply as a result of what can be best called strength of personality. We are so used to these inequalities of relationship that we often assume they are inevitable, the only way things can get done. For certain complex economic tasks hierarchical organization is indeed essential. But it is not necessary for many of the simpler things in day-to-day life. Insect communities appear to function very efficiently without discernible ranking. Man seldom seems to. This may be because he has instinctive proclivities to form relationships of a dominant or submissive kind and hence not only his daily life but also his social structure tends to take on this pattern.

In these rank and status relationships, man's behavior generally appears to reflect the proclivities postulated. (i) Challenges to an established position normally produce aggressive countering responses. Words like "insubordination" always smack of resentment. In all societies conflict and hostility are aroused when the established order is questioned. (ii) A faculty to submit readily, to accept leadership, to avoid conflict with those felt to be dominant also appears general in all human societies; were it not so social peace would indeed be a rarity. (iii) Affection, respect, and other manifestations of amicable behavior commonly coincide with rank and status relationships; many a military commander has the affection of his men, many a forceful husband is loved by his gentle submissive wife.

GROUP FORMATION IN SOCIETY

Human societies are made up of a large number of interlocking groups. The family, the town or village, the football team, the religious denomination, and particularly the occupational, professional, or trade group, are all essential features of the way we live. Among primitive peoples also anthropologists recognize that the understanding of any society presupposes an understanding of its social groups. There are the family units, then kinship groups, made up of related families, then lineage groups, clans or gens. A number of clans make a tribe and above that there may be groupings of tribes coming together for ceremonial, religious, or military reasons when the term nation is used. There are always other parallel social groups: magico-religious associations with membership qualifications, initiation ceremonies, and the like. There are many economically determined groups, hunting parties, canoe groups, etc., who continually work together and develop a sense of identity based on the use of technique or method. There are groupings of children of similar age "ganging up" for play just as do children in a modern city. The form of both modern and primitive so-

cieties is to be measured not only in the activity of individuals but in the complex overlapping of many constituent groups.

The ape group is not an amorphous conglomeration of individuals; it has a definite structure, the elements of which can be seen in the social groups of men. Blue-eyed men may be a "group" to a geneticist working on pigment inheritance but they do not make a social group. They have no sense of common identity. They do not have a hierarchy, nor bonds of amicable co-operation and friendliness. A social group in general has these things. It has continuity even though its members may have other group affiliations. Under stimulus of challenge, group conduct is ready to be called out. Let some government levy a tax upon blue-eyed men. Blue-eyed men will quickly form themselves into a social group with a common sense of identity, leaders will be chosen, others will follow loyally, friendliness and camaraderie will develop.

This tendency to identify oneself with certain other people and in that context to behave as a group may well flow from man's instinctive equipment. Human social groups, for all that their determining and distinguishing characteristics are historical and cultural, show features of the subhuman primate groups.

GROUP CHALLENGES AND HUMAN HISTORY

The idea that challenges play an important part in shaping the course of human affairs is in no way new. Toynbee in his "Study of History" (1939–54) makes considerable use of the idea. It seems to the present author, however, that it is essential to make a clear distinction between two different types of challenge. The first is the "direct challenge"; a challenge perceived as such by the group. Insofar as human social groups are compounded of the same instinctual ingredients as the ape group, such a challenge will increase group-cohesive behavior. The hostility of members of the group will be turned outward toward the source of challenge. Intragroup hostilities and conflicts will be diminished, the dominance scale within the group more readily accepted. The direct challenge, in short, integrates the group to repel assault.

The other sort is the "indirect challenge." This is a challenge that is not perceived as such at all. It is a challenge primarily of changing events, sometimes natural, more often technological and economical. Such changes insidiously alter the way in which the individuals within a social group work, and sooner or later this calls into question the established scales of power. In consequence there will be increasing challenges with overt conflict between individuals; there will be a diminution of amicable relationships and group loyalties. The effect of an indirect challenge upon a group is disruptive.

One may use the term "society" to signify the largest group with which its members have a sense of conscious identity. In practice this has usually been more or less synonymous with nation or state. Within each society there are constituent groups, often with opposing interests and with some degree of conflict between them. A direct challenge to a society as a whole—in practice this will mean military assault—will be integrative, and will minimize conflict between its constituent social groups. An indirect challenge to a society—normally by developing economic methods—will alter the balance of power between constituent social groups, particularly those that have an economic delineation, and will bring them into conflict. As a result, conflict, by challenge and counterchallenge, will strengthen and integrate the relevant social groups. Indirect challenge, whilst disruptive to a society as a whole, would often be integrative to constituent groups within.

By analyzing the sort of social groups that wax and wane through any given period of history and by describing the part played by direct and indirect challenge, one can, in fact, draw an intelligible picture. In the appendix two such periods, one covering the English revolutions of the 14th and 17th centuries, one covering the rise and fall of Rome, are given. Necessarily these descriptions are but thumbnail sketches: they grossly oversimplify, they entirely omit

many of the formative currents. But they are accurate. (Historians disagree with one another about what is and is not historically accurate. We can only achieve accuracy within those limits.) It is believed that an historical analysis of this kind shows quite clearly a recurring pattern of events which corresponds to the pattern of human proclivities suggested.

CONCLUSION

The idea that there are instinctive ingredients in man which tend to shape his history and his social structures is one that has not yet been much explored. Here it has been suggested that in the areas of dominance-submission and of group formation, comparative ethology may have something to offer the psychologist, sociologist, and historian. The value of an hypothesis, particularly a specious one like this, is in its heuristic potential. This one has the virtue of suggesting investigation in four directions:

1. In animal ethology: further study of the parameters of challenge, and its effect upon the group activity of the primates and lower mammals, is indicated.

2. In social psychology: the effect of direct and indirect challenge upon the sociometric patterns of small groups would appear worthy of investigation.

3. In clinical psychology: the possible role played by dominance and submission drives in some cases of neuroses could be considered. Some aspects of Adlerian teaching may be ready for reappraisal.

4. In economic and social history: specific periods can be examined to see whether they can legitimately be interpreted in the terms of group formation and challenge here suggested.

APPENDIX

The English revolutions

The most clearly drawn revolutionary conflict in English history was not that of the Cromwellian revolution, but that preceding the peasants' rebellion in 1381. Then there were only two groups involved, the poor peasants and landless laborers against the rich lords and landowners.

The whole of the fourteenth century had been an age of agricultural expansion. The old feudal manor was steadily changing in character, and estates, both large and small, were growing fast. Both wage labor and tenant farming were increasing. Merchant wealth was only just beginning, and national prosperity was in general still a matter of agricultural prosperity.

The agricultural expansion of the first half of the century was little reflected in the conditions of the peasants and laborers. They lived on a bare subsistence level. Suddenly all the changes taking place in the old feudal system were accelerated by the coming of the Black Death. Within a year it killed about a third of all the population. Overnight the economic situation was changed. The essential shortage was no longer of land or beasts or implements, but of labor. With the harvests rotting in the fields, the laborer suddenly found he could demand and get higher wages than ever before. The peasant found he could give a lesser service, and the bonded serf could leave his master and wander elsewhere, with but small chance of being caught, and every prospect of bettering himself. The base-line of poverty, and submission to the feudal master, previously accepted as inevitable, was now questioned. Society as a whole had been suddenly altered by the indirect challenge of a changing economy accelerated by the plague.

The feudal lords and landowners responded to what was a direct challenge from the laboring classes. They passed a whole series of laws designed to keep the worker on the land and make illegal any increase of wages. They tried long and hard to enforce these laws, often by the most violent methods. A great unrest disturbed the whole countryside.

In their turn, the peasants and laborers responded to the direct challenge of the landlords' oppressive legislation. For the first time a group organization grew up amongst them. They called it "The Great Society." At first it filled the role of a trade union, and arranged strikes for higher wages. With the success of these activities it grew stronger. As more working men and runaway serfs

moved around the countryside, the old isolation of agricultural life was undermined. As group feeling began to grow, and the Great Society became its organized expression, the teachings of Wycliffe and the Lollards began to be listened to as the peasants integrated into a self-conscious group under the influence of direct challenge.

How it was that the Great Society changed from a purely trade union body into a revolutionary instrument is not shown by the historical records of the time. But we know that the rebellion, when it came, was well organized, orderly, and devastatingly complete. In 1380 the government imposed a poll tax on every adult and it was this final direct challenge that precipitated the uprising. Almost simultaneously throughout the home counties the peasants rebelled. Within a few days London was occupied and the power of the land was in their hands.

The philosophy of such a peasant rebellion incorporates no notion of revolutionary government. The insurgents laid their demands before the captive king. When he promised to meet their grievances they disbanded and returned to their homes. Their grievances were not met. The king's soldiers and executioners saw to it.

Nevertheless the lords had been badly frightened. The poll tax was repealed and wages were allowed to rise. It is more than possible that the rebellion would later have recurred. But the situation began to change by new indirect challenges, again the result of changing economy. First there was a rapid expansion of the newer forms of land tenure, giving rise to a new class of small yeoman farmers sometimes employing labor themselves. Second, and more substantial, there was a rapidly increasing growth of merchant capitalism with the development of commodity production centered in the towns. New sectional groups grew up around these productive systems, and the harsh dichotomy of peasant and landlord was blurred.

The English revolution of 1640, in terms of economic antagonism, was a much less clear-cut affair. In the final analysis it replaced the old political system that had grown up under feudalism with a new form more in keeping with capitalism. But the complex tangle of events through which this happened calls for an understanding of social group formation.

In the first instance the issues of contention were quite well defined. Charles the First was following a policy aimed at increasing royal dominion. The heavy taxes he sought to raise had naturally borne heaviest upon the new class of merchant capitalists, whose assets were mobile and more taxable than those of the big landowners. They had also fallen heavily upon the smaller farmer and yeoman with only limited resources. His demands were direct challenges which strengthened and integrated both these groups. Their political power, such as it was, lay in the House of Commons. Their religious views, by and large, were puritanical, especially since a group of high churchmen under Laud surrounded and advised the King. Thus as a coalition arose in response to the King's challenges, its beliefs incorporated both parliamentarianism and Nonconformism.

Parliamentary refusal to vote the required taxes was a direct challenge to the King's status, and he responded by a series of hostile actions designed to weaken parliament. These had the effect of all direct challenges, integrating previously heterogeneous interests into a coherent group bitter in their opposition to him. His stubborn and inept handling of affairs precipitated a revolutionary situation with remarkable rapidity. He seems to have been quite devoid of that sense of statecraft which, in the Elizabethan era, had served the monarchy so well. Even during the Long Parliament, after the impeachment and execution of Strafford, he might still have counted on many in the Commons as his supporters. But he challenged even them by marching into the House with armed men summarily to arrest their leaders. But having been warned of the plot they had fled to the City, and the challenge touched off the revolution. The London trained bands rose in defense of Parliament. Charles fled to York and to the protection of his armies in the agricultural provinces.

Unlike the peasant rebellion, the insurgent groups now were well-versed in the philosophy of government, and parliamen-

tary institutions were at once able to take over the responsibilities of the state. The revolutionary coup d'etat as such was over within a few days.

The King's armies in the provinces had military experience and tradition. For some time he was not only able to resist defeat, but made headway against the parliamentary forces, with their none-too-adept generals. It was at this time that Cromwell rose to power.

Son of a landowner, a puritan, a man of great integrity, Cromwell was the military genius of the era. He formed a new force. Well-disciplined, regularly paid, politically indoctrinated and thoroughly trained, and with a self-conscious sense of group identity, this New Model Army soon swept the Royalists before it and turned the tide of the war. In 1646 the King surrendered at Newcastle, and the revolution was won.

A revolutionary coalition is held together by direct challenge. Victory itself is an indirect challenge, breaking apart the groups that make it up in a post-revolutionary struggle for power. So it was with the English revolution.

The Parliamentary forces were a coalition of two main groups. The Presbyterians were rich merchants, moderate-sized landowners and country squires, and were supported usually by the people of the towns. The Independents were mainly the small farmers and the artisans and craftsmen from the provinces. The Presbyterians were in a majority in the Commons, but the New Model Army was mainly Independent.

With the end of the civil war rancor and conflict began between the two wings. The captive Charles was playing one off against the other, whilst Cromwell was trying to keep the two groups together. There was a series of direct challenges by the Presbyterians in the House, including the passage of acts to disband the army with its arrears of pay still outstanding. This, of course, served only to integrate the army around the Independents. When finally a Presbyterian plan to reconstitute the Royalist Army was discovered, a second civil war (1648) broke out. It was short and violent and it ended in victory for the New Model Army and the Independent Party. A body of troops were marched into Parliament and the Presbyterians expelled forthwith. The Independents, the 'Rump,' took over complete control.

But within a New Model Army itself there were also group divisions. The ordinary soldier, embracing the radical policies of the Levellers, was deeply distrustful of the 'Grandees,' the richer officer caste of which Cromwell was a member. In the period between the two civil wars this conflict had blazed out with some ferocity, only to die down under the direct challenge of the Presbyterians. But victory was again an indirect challenge, and the struggle began once more. The suppression of all differences and insubordination, vigorously prosecuted by Cromwell, really brought the Commonwealth to its end. One by one the social groups that had brought it into being were challenged and antagonized, and finally it rested only upon a purged and weakened army and upon the reputation of Cromwell himself. Soon after his death the monarchy was restored. Quietly, and almost without opposition, Charles II was invited to return, invited to be a constitutional monarch whose power was to rest on a Parliament reconstituted as the political apparatus of the bourgeoisie landowners and merchant capitalists.

In the end what had happened was that the old monarchy, born of feudal conditions, striving for complete dominance against the growing merchant capitalism, had been replaced by a new monarchy, supported by and supporting the new bourgeoisie. The revolution had completed the change from feudalism to capitalism. But the tale of the struggle is one of growing and interacting social groups, integrating under direct challenge and breaking down under indirect challenge.

The rise and fall of Rome

The township of Rome on a ford of the Tiber grew up about the beginning of the eighth century B.C. At that time the northern part of Italy was the site of a thriving Etruscan society, whose hegemony spread down to the plain of Latium. The country south of the river was the home of some

primitive agricultural societies, organized around a dozen fortified townships. It seems probable that in its early days the towns had a local king or lord who governed with a council of elders. The Latin tribes all shared some sense of identity, spoke a common language, and had certain common religious functions, though they were never known to operate as a single governmental unit.

Rome, being on the first ford of the Tiber, and being navigable to ocean-going ships, soon developed in importance. It fell under the Etruscans (the Tarquins) who took the kingship and ruled it as part of their empire. Although this century of the Tarquin dynasty was tremendously prosperous, and was attended by much building and a great increase of trade and by the accumulation of capital assets, the foreign rule was resented. The native Romans did not identify with their masters. The regime became more tyrannous. Both the Senate, compounded of the richer Romans, and the ordinary Roman citizens grouped together under the challenge of foreign rule. A revolutionary situation developed steadily, but could not achieve success until the Etruscan military power was defeated elsewhere. In 508 B.C. this happened, and the revolution threw out the monarchy. The most powerful element of the Roman coalition, the Senate, became the dominant group in post-revolutionary Rome.

The conclusion of the old conflict was an indirect challenge. From it a new struggle began. This was the conflict of the Patricians and the Plebeians. The Patrician families were the nobility, and the governing Senate was made up of their elders. They possessed the merchant wealth and were the big landowners. The Plebeians were the free Roman citizens, that is, the smaller landowners, the craftsmen and laborers, and the tenants of the larger patrician estates. The two groups were clearly demarcated, intermarriage was forbidden, and there were separate religious observances as well as differences in political status. Both groups were Romans. They felt an identity together. They distinguished themselves from the slaves and non-Roman laborers within, and from the other Latin tribes and other societies without. Both served together in the Roman Army, contributing their service and their taxes in accordance with their economic and social rating.

There was no great period of peace after the overthrow of the Tarquins. Not only had the struggle for power between Plebeian and Patrician begun almost immediately, but Rome also became embroiled in repeated wars. For three centuries Rome came into conflict with her every neighbor. Amicable relationships were never anything more than shortlived opportunist military alliances. The Romans fought the other Latin cities. They fought the Etruscans and the Gauls. They fought the Aegui and the Volsci. They fought the Samnites and the southern Italians. They fought the Greeks. Sooner or later they won each war, and by 270 B.C. were dominant over all Italy. These wars were not part of any consistent plan of expansion. There was no Alexander or Caesar following a deliberate policy of military conquest. They arose because that is the way that men conduct their intergroup affairs. The mere fact that a society has a powerful neighbor is sufficient to act as a challenge and lead to conflict.

Each war, calling upon the men of both classes to fight in the legions, was a direct challenge and was integrative. But each victory was an indirect challenge. The poorer citizen could not maintain his land whilst under arms; it was mortgaged and finally expropriated by the big Patrician landowner. The struggle between Patricians and Plebeians flared up after every victory. In 494 B.C. after the Latin war, the returning legionnaires marched out of the town, and threatened to form a new city higher up the Tiber. The Patricians yielded, and the Plebeians got their own tribunes and officers. A few years later, again to obtain legal rights, the Plebeians once more withdrew from Rome, and again the Patricians yielded. Repeatedly, over two centuries, the Plebeians, rallying behind some dominant ambitious leader, came into conflict with the Senate. Revolution seemed often to be close.

But the conflict never developed fully. The growing power and scope of Rome during this period brought changes which modified and finally replaced the struggle altogether.

At root these changes were economic. There was the tribute from conquests, with the possibility of land and wealth in the conquered territories, and there was a growing capitalism of trade and production displacing the older form of landed wealth. Many of the Patrician families fell upon hard times. Some Plebeians became wealthy, even very wealthy. The political concessions wrung out by militant Plebeian action gave opportunities for dominant Plebeians to gain power within the constitutional framework. The laws against intermarriage were annulled in 445 B.C. Steadily office after office became open to Plebeians, and finally even the Senate itself. These changes were indirect challenges to both groups, but particularly to the Plebeians. The economic differences which had distinguished the Plebeians from the Patricians became blurred. By the time of the First Punic War, 264 B.C., the Patrician-Plebeian conflict was dormant.

The three Punic Wars against the prosperous empire of Carthage (264, 218, and 149 B.C.) stand as examples of the spontaneous purposeless nature of war. In material terms there was absolutely no need for these struggles. There were territories in Europe open to Roman expansion, and territories in Spain and Africa open to the Carthaginians. Their trading needs could, with a little good will, have been adjusted to mutual satisfaction. They were two peoples who, in terms of common sense, could have coexisted profitably. But they were two nations, each formidable, and so they must constitute a direct challenge to the other. The wars brought death and pillage, suffering and disease. They ended at last with the killing of every single Carthaginian and the total destruction of the city.

Meanwhile in Rome the conflict between the old power of the Senate and the dispossessed Romans also mounted. This struggle between the 'Optimates' and the 'Populares' lasted for half a century. It might well have come to revolutionary intensity had not both groups been repeatedly subject to influences which cut across the group lines dividing the contenders.

The non-Roman Italians, excluded from government and franchise and overtaxed to subsidize Rome, finally rebelled. The 'Social Wars' began in 90 B.C., and Roman legions were marched up and down the country, subjugating and pillaging. The direct challenge was itself integrative to Rome. One Roman army was commanded by Marius, a consul who was regarded as a favorite of the Populares, and another by Sulla, a general of the Senatorial Optimates. Together they suppressed the rebellion.

Foreign wars, also by constituting direct challenges, had an integrative effect. The century was one of enormous imperial expansion, no longer haphazard but planned to give field for imperialist colonial expansion comparable in its economic background to that of nineteenth century Europe. The Mithridatic War in particular, erupting during the Social War, had an integrative effect on the whole of the Italian peninsula. Roman citizenship was at once granted to all Italians. The Social War petered out, whilst the Mithridatic War was fought and won

The Punic Wars, by keeping the legions under arms for many years and by sending them so long overseas, had created the professional Roman soldier. Drawing pay and seeking loot, he had abandoned his sense of identity with the Roman civil population, and formed a new self-conscious social group. The professional legionnaires, accepting a scale of dominance up to the commanding general, were soon destined to become more and more a social force in the conflicts of Roman history.

By the middle of the first century B.C. the social schism was deepening rapidly. The poor and dispossessed of Italy, now all Roman citizens, had begun to polarize into a distinct group in conflict with the dominant capitalist and senatorial coalition. But the new social group had entered Roman politics. After the Social War was over, the two principal military leaders, Marius and Sulla, had come into conflict, and on different occasions had used their legions to massacre their political opponents. From that time on the success of every struggle for power in the Senate, and every struggle between the dominant groups and the dispossessed, had been dependent upon alliance with military power. Inevitably the time came when the

generals went into business on their own account. In 49 B.C. in Thessaly, Roman legion fought Roman legion; Julius Caesar defeated Pompey, and on returning to Rome with his army, was made dictator. After his assassination Republican Rome came to its end in a long struggle for the emperorship; also a struggle which turned upon military power.

The dictatorship of the Emperors at first constituted a mediating force between the two classes. But as the indirect challenge of growing economic differentiation, growing riches and growing poverty mounted, the dictatorship, to keep the peace, was obliged to throw its power on the side of the economically dominant groups. This was a direct challenge to the dispossessed groups, and the dictatorship became steadily more resented and more tyrannical. Under a series of oppressive challenges, the Roman populace, the Italian peasants, the landless laborers, the freed slaves, and finally even the underpaid soldiery came to coalesce together in discontent. Rome was at last facing revolution.

But the autocratic power of the emperors was not easily overthrown, and the conflict mounted over many years until finally even direct challenges from outside ceased to have an integrative effect. The emperors marched their legions from one area of barbarian invasion to the next. But the people, particularly the provincial people, were apathetic. They no longer identified themselves with Rome. Indeed, often they welcomed the barbarian as their relief from Roman tyranny. Rome met defeat after defeat from outside. Column by column the whole edifice of the Roman Military Empire collapsed.

This outline of Roman history, though discerning readers may feel there are inaccuracies of emphasis here and there, does show how the growth and decline of a society can be seen in terms of social groups.

The internal history of a society is in large part the history of social groups, many of them delineated by the economic status of their members. Each group integrates as it meets direct challenge and disrupts under the indirect challenge of change. The growth of new groups, often as a result of economic development, alters the issues of social conflict and gives dynamic impulse to a society throughout much of its history.

REFERENCES

Carpenter, C. R. A field study of the behaviour and social relations of howling monkeys. *J. Exp. Psych.*, 1934, 31, 272–292.

Carpenter, C. R. A field study of the behaviour and social relations of the gibbon. *Comp. Psychol. Monogr.*, 1940, 16, 5.

Koehler, Wolfgang. *Mentality of apes.* London: Routledge, 1925.

McDougall, W. *An introduction to social psychology.* London: Methuen & Co., 1908.

Toynbee, A. J. *A study of history.* Oxford: Oxford University Press, 1939–54.

Zuckerman, S. *Social life of monkeys and apes.* London: Kegan Paul, Trench, Trubner & Co., Ltd., 1932.

(Manuscript received August 4, 1960.)

Cromwell was about to ravage all Christendom; the royal family was undone, and his own forever established, save for a little grain of sand which formed in his ureter. Rome herself was trembling under him; but this small piece of gravel having formed there, he is dead, his family cast down, all is peaceful, and the king is restored.

PASCAL, *Pensées*

The construction of an interval scale by equisection methods requires only that the subject partition a segment of a "physical" continuum into "subjectively equal" intervals with respect to the psychological attribute which is being scaled. Although many investigators have used these methods to construct scales, interest has focused primarily on problems connected with the psychophysical law. Here attention is devoted to tests of the validity of the constructed scale by providing an analysis that makes explicit those assumptions implicit in the use of equisection scaling methods.

A MODEL FOR EQUISECTION SCALING

*by Robert F. Fagot**

University of Oregon

INTRODUCTION

THE construction of an interval scale by equisection methods requires only that the subject partition a segment of a physical continuum into "subjectively equal" intervals with respect to the psychological attribute.[1] The technique has been applied primarily in the area of sensation, but it is applicable in any area in which stimuli are available either continuously or in sufficiently small discrete steps to guarantee that stimuli which lie at equal intervals on the subjective continuum are available to the subject. Torgerson (1958, pp. 117–131) provides a good description of the methods and an analysis of the major problems. An early history of empirical results can be found in Boring (1942, pp. 42–43), and Titchener (1905, pp. 210–232).

Although many investigators have used these methods to *construct* scales, interest has focused primarily on problems connected with the psychophysical law (see Stevens, 1957), and little attention has been devoted to *tests* of the "validity" of the constructed scales (tests of internal consistency). The significance of psychophysical laws based on psychological scales which have not been validated is dubious at best.

The intent of this paper is to provide a theoretical analysis which makes explicit those assumptions implicit in the use of the equisection scaling methods, and which permits the derivation of all testable consequences of the scaling assumptions. It will be seen that the emphasis here is on tests of internal consistency and formal problems connected with the finite set of stimuli used to construct the scale, rather than on the traditional problems of psychophysical laws.

THE FORMAL MODEL

S is the set of stimuli with elements $\{x, y, z, \cdots\cdots\}$. There is no limit on the admissible number of stimuli, and in most applications S will be an infinite set, since a physical continuum will usually be available. If stimuli differ with respect to one variable (e.g., tone frequency), S can be represented by a segment of the real line. $S' = \{x_1, x_2, \cdots, x_n\}$ is a finite subset of S. The n elements of S' are the n stimuli which partition the closed interval $\{x_1, x_n\}$ into $n - 1$ subjectively equal intervals. Thus the stimuli in S must be available either continuously or in small discrete steps to guarantee that stimuli which lie at equal intervals along the "subjective continuum" will be available to the subject. We shall see that S' must have at least three members to *construct* a scale, and at least four mem-

* This research was supported by a grant from the National Science Foundation. I am indebted to E. W. Adams for valuable critical comments and suggestions.

[1] Equisection scaling of subjective values is formally analogous to paired comparisons between even-chance bets (Suppes & Walsh, 1959).

bers to *test* the scale. P is a binary relation on the set S, and is interpreted such that xPy if and only if the subjective value of x is judged *greater than* that of y. If the stimuli are tones which differ only in frequency, then xPy if and only if the pitch of x is judged greater than the pitch of y. Actually, in applications, observations of P relations are usually neglected, but assumptions about P cannot be omitted from the formal model. E, Q, and R, are quaternary relations on S. E is an equivalence relation; $xyEzw$ if and only if the difference in subjective value between x and y is *equal to* the difference between z and w. For tones differing only in frequency, $xyEzw$ implies that the difference in pitch between x and y is equal to the difference in pitch between z and w. If y is the stimulus which bisects the subjective distance between two stimuli x and z, then $xyEyz$. $xyQzw$ holds if and only if the difference in subjective value between x and y is *greater than* the difference between z and w. For reasons of formal simplicity, certain of the assumptions are formulated in terms of the quaternary relation R, rather than in terms of both E and Q. R has the numerical interpretation *greater-than-or-equal-to*. E and Q are defined in terms of R as follows:

Definition 1 (E): $xyEzw$ if and only if $xyRzw$ and $zwRxy$.

Definition 2 (Q): $xyQzw$ if and only if not $zwRxy$.

It should be clear that all observations of quaternary relations will be in terms of E and Q, and not in terms of R. However, the necessary "theorems" expressed in terms of E and Q follow directly from the fundamental assumptions on R together with the above definitions. For example, from the assumption that R is transitive and the definition of E above, we may derive the transitivity of E.

It should be noted that the relations are customarily *interpreted* in such a way that each relation is dependent upon more than one judgment. Thus $xyEyz$ might mean that y is the mean or median of a distribution of judgments, each judgment being a report by the subject locating a stimulus which he perceives as bisecting the distance between x and z.

We begin the construction of the model with the following condition, which asserts the existence of the scaling function u.

Condition 1

There exists a real-valued function u for the set S and relations P and R which satisfies the following conditions for every x, y, z, and w, in S:

(a) xPy if and only if $u(x) > u(y)$, and if $k \neq y$, then $u(x) \neq u(y)$.

(b) $xyRzw$ if and only if $u(x) - u(y) \geqq u(z) - u(w)$.

Note that condition (a) rules out the possibility that two different stimuli may be judged equal in subjective value. However undesirable this may seem, the exclusion is consistent with the assumptions implicit in the use of the equisection methods of scaling.

It is clear that Condition 1 as stated can never be completely tested, since S is infinite. But for any finite subset of S, such as S', a decision procedure can be given for testing unequivocally whether or not u exists (for the set S' and relations P, Q, and E). Corresponding to each observation is an equality or inequality in u, and hence the function u exists if and only if the resulting set of equalities and inequalities has a solution. The fact that u exists for S', however, does not in general guarantee that u is unique up to a linear transformation. Thus if the experimenter selects n stimuli at random and has subjects compare all possible intervals, then all of the judgments might be of Q relations (and will therefore generate inequalities only), in which case the interval scale could not be constructed, since for such a construction special E relations (and corresponding equations) are required. For example, if we have the set of inequalities $u(x_3) - u(x_1) > u(x_3) - u(x_2) > u(x_2) - u(x_1) > 0$, and fix origin and unit by putting $u(x_3) = 1$, $u(x_1) = 0$, then *any* proper fraction smaller than $\frac{1}{2}$ can be assigned as $u(x_2)$. The special conditions peculiar to equisection scaling do, however, guarantee that the required E relations will be observed. But before considering these conditions of equisection scaling, the notion of the "construction" of the scale needs some clarification. *By the CONSTRUCTION of*

an interval scale for the set S' will be meant that the assignment of $(n - 2)$ scale values is uniquely determined from the observations and Condition 1, given the arbitrary selection of the unit and origin of the scale.

Thus the observations (1) $x_3x_2Ex_2x_1$ and (2) $x_4x_3Ex_3x_2$ imply, by Condition 1: (1) $u(x_3) - u(x_2) = u(x_2) - u(x_1)$, and (2) $u(x_4) - u(x_3) = u(x_3) - u(x_2)$. These $(n - 2 = 2)$ equations are sufficient to construct the scale for the four stimuli. Assigning $u(x_1) = 1$ and $u(x_2) = 2$ (arbitrary selection of origin and unit), the values of the two remaining stimuli are uniquely determined as: $u(x_3) = 3$ and $u(x_4) = 4$. Thus it is always possible to construct $(n - 2)$ *independent* equations in $(n - 2)$ unknowns for which therefore a unique solution exists. Any such set of $(n - 2)$ independent equations or the corresponding set of $(n - 2)$ independent E relations will be referred to as a *construction set*. The intent is to provide two conditions, each of which guarantees the existence of a construction set. One value of such conditions lies in the fact that each stipulates the minimum number of observations necessary to construct the scale, and hence maximizes the number of observations available for a *test* of the scale, since all observations not used in constructing the scale are available for testing it. Before presenting these two conditions, we shall consider two types of judgments, i.e., comparisons of intervals, which are involved in the determination of a construction set.

In both these types of judgments the subject is presented with two stimuli x and z which are ordered xPz. If his task is to find a stimulus y such that xPy, yPz, and $xyEyz$, then this type judgment is called a *bisection judgment*, which is consistent with conventional terminology. If his task is to find a stimulus w such that zPw and $xzEzw$, or wPx and $wxExz$, then such judgments are called *adjacency judgments*. The exclusive use of either of these two types of judgments can lead to a construction set. The sole use of adjacency judgments is sufficient to derive a construction set satisfying Condition 2, and the sole use of bisection judgments is sufficient to derive a construction set satisfying Condition 3. Condition 2 follows.

Condition 2 (Adjacency)

(a) $x_{i+1}Px_i$

for $i = 1, 2, \cdots\cdots, n - 1$

(b) $x_{i+1}x_iEx_ix_{i-1}$

for $i = 2, 3, \cdots\cdots, n - 1$.

Condition 2 imposes the desired ordering and "equal-spacing" of the n stimuli in S'. In the example above, for $n = 4$, observations (1) and (2) provide the E relations required by (b) and hence partition the interval $u(x_4) - u(x_1)$ into three subjectively equal intervals. We may imagine the subject, with his subjective ruler in mind, successively "laying off" equal intervals, the first interval defined by the experimenter (presentation of x_1 and x_2 or x_n and x_{n-1}), and each successive interval constructed by the subject equal to the immediately adjacent one, the process being continued for $(n - 2)$ "steps." These $(n - 2)$ observations provide a construction set, and the experimenter may now perform a test of internal consistency by testing the remaining observations, all of which are implications of the constructed scale. For the example above, $(n = 4)$, only two such E relations remain, (3) $x_4x_3Ex_2x_1$, and (4) $x_4x_2Ex_3x_1$. These two observations can be predicted from the constructed scale by Condition 1, and provide the test of internal consistency for S'. For larger n there will, of course, be a larger number of such implications of the scale.

We have seen that the Adjacency Condition begins with the experimenter defining a unit interval, and the judgments of the subject provide the remaining $(n - 2)$ unit intervals, which partition the total interval into $(n - 1)$ subjectively equal unit intervals. The bisection method, on the other hand, begins with the total interval; i.e., the subject is presented with stimuli x_1 and x_n and by successive bisections partitions this total interval, $u(x_n) - u(x_1)$, into the required $(n - 1)$ equal (unit) intervals. At the first "stage" in the bisection process, the subject bisects the total interval; i.e.,

partitions this interval into two subjectively equal intervals. At the second stage, each of these *two* intervals is in turn bisected, providing four intervals; at the third stage, each of these four intervals is bisected, and the process is continued for p stages where

$$n = 2^p + 1. \tag{2}$$

Unfortunately, in spite of the fact that one can give a simple verbal description of the bisection process, a mathematical formulation of this method of constructing a scale is unavoidably complicated. A sketchy outline of the reasoning leading to the formulation of such a condition is given below.

At the 0th stage we have x_1 and x_n ; this corresponds to the presentation of the "terminal" stimuli which define the total interval. At the 1st stage the process generates x_1 , x_i , and x_n corresponding to the relation $x_n x_i E x_i x_1$, where n is given by (2) and $i = 2^{p-1} + 1$. 1, i, and n can be written: $1 = 0 \cdot 2^{p-1} + 1$, $i = 1 \cdot 2^{p-1} + 1$, and $n = 2 \cdot 2^{p-1} + 1$. This way of writing the subscripts provides the key which enables one to stipulate precisely which intervals to bisect at each stage of the process, and this reasoning leads to the following formulation of the bisection condition.

Condition 3 (Bisection)

(a) $x_n P x_1$

(b) $x_h x_i E x_i x_j$

where

$$h = (k + 1)2^{p-r+1} + 1$$
$$j = (k)2^{p-r+1} + 1$$
$$i = (h + j)/2 = (2k + 1)2^{p-r} + 1$$

for $r = 1, 2, \cdots\cdots, p$

and $k = 0, 1, 2, \cdots\cdots, (2^{r-1} - 1)$.

A better understanding of this condition can be gained if we examine it for $p = 2$; then $n = 5$. We have three cases only to consider: (1) $r = 1$, $k = 0$, (2) $r = 2$, $k = 0$, and (3) $r = 2$, $k = 1$, which exhausts the implications of Condition 3 for $n = 5$. (1) implies $h = 5$, $i = 3$, and $j = 1$; i.e., $x_5 x_3 E x_3 x_1$, which corresponds to the first bisection of the total interval. (2) and (3) correspond to the second stage of the two-stage process in which each of the two intervals generated by the first bisection are in turn bisected. (2) implies $h = 3$, $i = 2$, and $j = 1$; i.e., $x_3 x_2 E x_2 x_1$, which corresponds to the bisection of the interval $u(x_3) - u(x_1)$. Finally, (3) implies $x_5 x_4 E x_4 x_3$, which corresponds to the bisection of the interval $u(x_5) - u(x_3)$. This completes the process since there are only two stages, and the three bisections have partitioned the interval $u(x_5) - u(x_1)$ into four equal intervals. From formula (2) we see that Condition 3 holds only for $n = 3, 5, 9, 17, \cdots\cdots$.

In the example presented in illustration of Condition 2, two adjacency judgments were sufficient to construct the scale, and the remaining possible E relations (3) $x_4 x_3 E x_2 x_1$ and (4) $x_4 x_2 E x_3 x_1$ were available as a test of internal consistency. The intervals used in (3) are nonadjacent, or separated on the scale, and E relations of this form will be called *separated* E relations, or separated judgments. E relations of form (4) will be called *overlapping* E relations or judgments; the corresponding intervals $u(x_4) - u(x_2)$ and $u(x_3) - u(x_1)$ overlap; i.e., these two intervals have the interval $u(x_3) - u(x_2)$ in common. Separated and overlapping E relations are logically related since Condition 1 implies $x_4 x_3 E x_2 x_1$ if and only if $x_4 x_2 E x_3 x_1$. As Conditions 2 and 3 demonstrate, adjacency or bisection judgments are sufficient to *construct* the scale, but for a *test* of the scale, separated and overlapping judgments are necessary.

Conditions 1 and 2 are the axioms for the *finite adjacency model* and Conditions 1 and 3 are the axioms for the *finite bisection model.* The stipulation "finite" results from the fact that the axioms permit only the assignment of scale values to the finite subset S' and not to the *infinite* set S, the latter requiring the further assumption of the *existence of a psychophysical function.* The "infinite" models will not be dealt with in this paper, but the problem will be discussed briefly in the concluding section.

Each of the finite models provides a different method of constructing the scale, and satisfaction of the model guarantees the existence of an interval scale. It will be recalled that a decision procedure was men-

tioned above for determining whether or not the function u exists for the set S'. It was shown that the fact that u exists does not in general guarantee that u is unique up to a linear transformation. However, if u exists *and* either Condition 2 *or* 3 is satisfied for S', i.e., if the adjacency model or the bisection model is satisfied, then the assignment of numbers to S' is unique up to a linear transformation.

Having constructed a scale by means of Condition 2 or 3, the next step is to perform the test of internal consistency. Although such a test is obvious from consideration of Condition 1, it is in order to state it formally as follows.

Test of internal consistency for S'

For all $h, i, j, k = 1, 2, \cdots\cdots, n$,
(i) $x_h P x_i$ if and only if $h > i$;
(ii) $x_h x_i Q x_j x_k$ if and only if $h - i > j - k$; and
(iii) $x_h x_i E x_j x_k$ if and only if $h - i = j - k$.

The subscripts, of course, give directly the scale values of the stimuli; i.e., for all x_i in S', $u(x_i) = i$. To illustrate the test of internal consistency, and to compare the use of Conditions 2 and 3 in constructing the scale, the case of $n = 5$ is examined.

If the scale is constructed by means of Condition 2, then the observations (1) $x_3x_2Ex_2x_1$, (2) $x_4x_3Ex_3x_2$, and (3) $x_5x_4Ex_4x_3$ provide the information necessary to assign scale values to the five stimuli. If (iii) of the test of internal consistency is to be satisfied, the following observations must be confirmed, since in each case $h - i = j - k$: (4) $x_5x_4Ex_3x_2$, (5) $x_5x_4Ex_2x_1$, (6) $x_4x_3Ex_2x_1$, (7) $x_4x_2Ex_3x_1$, (8) $x_4x_2Ex_5x_3$, (9) $x_5x_3Ex_3x_1$, and (10) $x_5x_2Ex_4x_1$.

If Condition 3 (Bisection) had been employed, the three observations used to construct the scale would have been (1), (3), and (9). Note, however, that if only bisection judgments are allowed the subject, then observation (2) is the only one of the remaining seven observations which can be tested for consistency. We have, then, the important conclusion that *if bisection is the only type judgment permitted, then although the scale can be constructed, a complete test of the interval scale hypothesis cannot be made.* Of course, the bisection model does *not* demand the sole use of bisection judgments, although such judgments are sufficient for the satisfaction of Condition 3. Once the minimum $(n - 2)$ bisections provide the construction set, the model imposes no restrictions on admissible types of E relations, and hence one can use bisection judgments to construct the scale and nonbisection judgments to test it.

There are, of course, certain P and Q relations which are implications of the scale, but in the usual experimental situation it would seem to be a waste of time to check these observations for S'. For example, the most difficult discrimination required of the subject involving a Q relation would be of the form $x_5x_3Qx_3x_2$ where the first interval is twice as large as the second. However, evidence is lacking on this point, and perhaps a minimum of such observations would be the wise procedure.

Thus the assignment of scale values to the stimuli in S' (by the construction of u) provides a direct means of making a test of internal consistency to determine if u exists for S'. Condition 2 or 3 guarantees that if u exists it is unique up to a linear transformation. If Condition 2 or 3 is satisfied, the *test of internal consistency for S'* provides a set of necessary and sufficient conditions for the existence of an interval scale for the specific stimuli involved in the construction of the scale.

SPECIAL "TESTABLE CONSEQUENCES" OF CONDITION 1

One interesting problem which arises in connection with Condition 1 concerns the possibility of deriving from this hypothesis "testable consequences" formulated in terms of observable relations. These consequences would be "theorems" derived from the basic assumption of the existence of u. For example, one obvious consequence is that the relation E is *transitive*. Thus in the last example in the previous section, observations (1) and (2) (used to construct the scale) imply (6) (an observation not needed to construct the scale) by the transitivity of E. The possibility then arises of using

such consequences to predict the outcome of comparisons of intervals not involved in the construction of the scale *without first constructing u*. The value of such consequences lies partly in the fact that if an interval scale does not exist (the model is disconfirmed), it becomes possible to state in precisely what way disconfirmation occurred. Thus it would be much more informative to know, for example, that E was not transitive rather than simply that the model was disconfirmed. For one thing, if the latter (general disconfirmation) one would still not know whether or not a weaker ordered metric scale existed, but knowing that E is not transitive immediately eliminates the possibility of such a scale existing.

Condition 4, which follows, presents six such testable consequences, each of which is a necessary consequence of Condition 1.

Condition 4 (Testable consequences of Condition 1)

For all x, y, z, w, u, and v in S:

4.1 Exactly one of the following holds: $x = y$, xPy, yPx.

4.2 If xPy and yPz then xPz.

4.3 If xPy and yPz then $xzQxy$ and $xzQyz$.

4.4 $xyRzw$ or $zwRxy$.

4.5 If $xyRzw$ and $zwRuw$ then $xyRuw$.

4.6 If $xyRzw$ then $xzRyw$ and $wzRyx$.

These consequences are formally similar to a set of axioms presented by the author in relation to a model for ordered metric scaling (Fagot, 1959), and to some of the axioms provided by Suppes (1957, p. 269) for "finite equal difference structures." Reference to the first of these sources may be made for a detailed discussion of their intuitive significance, but it might at least be pointed out here that 4.6 relates separated and overlapping judgments. Thus for $xPyPzPw$ we have from 4.6: $xyEzw$ if and only if $xzEyw$; where the first E relation is a separated judgment and the second is an overlapping judgment.

The first two consequences are the axioms for an *ordinal* model, whereas the first five are the axioms for an *ordered metric* model (Fagot, 1959, pp. 160–162). All six consequences do not guarantee an interval scale unless Condition 2 or 3 is satisfied, as stated in the following theorem:

Sufficiency Theorem for Condition 4: Conditions 2 and 4 *or* Conditions 3 and 4 are *sufficient* to guarantee the existence of a function u satisfying Condition 1, and the *uniqueness* up to a *linear transformation* of the numbers assigned to S'.

We have already seen that, given a construction set, all remaining E relations (as well as Q relations) can be predicted by Condition 1. The above theorem guarantees that such predictions can also be made by Condition 4.

The formulation of these consequences serves to make explicit the formal similarity of an ordered metric scale and an "equisection" interval scale. It is seen that, within the context of models involving direct comparisons of intervals, it is the addition of consequence 4.6 and Condition 2 or 3 which transforms the ordered metric scale to the "stronger" interval scale. And since Conditions 2 and 3 are always satisfied by virtue of the nature of the scaling procedure, it is 4.6 which becomes the crucial condition (a condition which demands of the subject overlapping judgments).

PREVIOUS EMPIRICAL STUDIES

The Bisection Condition has been used frequently to construct scales in the area of sensation (Gage, 1934; Newman, Volkmann, & Stevens, 1937), but only one study has come to the attention of the author in which the Adjacency Condition was used (Judd, 1933). Münsterberg (1892) and Pratt (1938) report studies in which separated judgments were required of subjects, but the author knows of no studies in which overlapping judgments were required, although such judgments are necessary for a test of the scale. Gage (1934) constructed scales for brightness and loudness by the method of bisection, made the one test possible using bisection judgments, and reported disconfirmation for each scale. Newman et al. (1937) repeated Gage's study on loudness, reporting more satisfactory results, and concluded that consistent results would be obtained if constant errors could be minimized. These two studies constitute only

partial tests and no other studies have been reported of tests of internal consistency for equisection scales.

"INFINITE" MODELS

The analysis presented in this paper has been concerned with finite equisection models; i.e., with problems related to the construction of the scale for the finite set S', and the tests of internal consistency for the constructed scale. Given a constructed scale, one may assume the existence of a real-valued monotonic increasing function relating the physical continuum and the subjective continuum—the psychophysical function. The problem, which has been given considerable attention (Stevens, 1957), is then to determine the functional relation between the two continua, i.e., to discriminate between possible psychophysical laws. The existence of such a function also permits the assignment of scale values to the infinite set S, and admits the possibility of comparing two scales constructed with different sets of stimuli, provided the range of the two scales overlap on the physical continuum. An analysis of these problems (essentially the construction of "infinite" models) will be reserved for a later paper, wherein the problem of discriminating between possible psychophysical laws will receive special attention.

REFERENCES

Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton Century, 1942.

Fagot, R. F. A model for ordered metric scaling by comparison of intervals. *Psychometrika,* 1959, 24, 157–168.

Gage, F. H. The measurability of auditory sensations. *Proc. Roy. Soc.,* 1934, 116b, 103–119.

Judd, D. B. Saturation scale for yellow colors. *J. Opt. Soc. Amer.,* 1933, 23, 35–40.

Münsterberg, H. Vergleichung von tondistanzen. *Beiträge zur exp. Psycholg.,* 1892, 4, 147–177.

Newman, E., Volkmann, J., & Stevens, S. S. On the method of bisection and its relation to a loudness scale. *Amer. J. Psychol.,* 1937, 49, 134–137.

Pratt, C. C. Bisection of tonal intervals larger than an octave. *J. Exp. Psychol.,* 1938, 11, 17–26.

Stevens, S. S., On the psychophysical law. *Psychol. Rev.,* 1957, 64, 153–181.

Suppes, P. *Introduction to logic.* Princeton: Van Nostrand, 1957.

Suppes, P., & Walsh, K. A non-linear model for the experimental measurement of utility. *Behav. Sci.,* 1959, 4, 204–211.

Titchener, E. B. *Experimental psychology,* Vol. II. New York: Macmillan, 1905.

Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

(Manuscript received July 5, 1960.)

Nothing irritates a student of culture more than to have the methods of the exact sciences flaunted in his face as a salutary antidote to his own supposedly slipshod methods. He feels that he deals with an entirely different order of phenomena, that direct comparison between the two groups of disciplines is to be ruled out of court.

EDWARD SAPIR, "Do We Need a 'Superorganic'?"

The great variety of ills that man is heir to and the tremendous number of symptoms associated with such diseases make the task of medical diagnosis more speculative than scientific. There is an expanding interest in the possibilities of using computers to help make medical diagnoses more quickly and accurately, but here again, the vastness of the task complicates the establishment of a general quantitative diagnostic procedure. Suggested here are models which simplify this procedure into a sequence of decision processes of manageable size for a statistical approach. Perhaps automatic data processing can make an initial inroad into the area of medical diagnosis through the accumulation and continuous revision of base rates appropriate to the various levels in the diagnostic decision process.

MODELS FOR MEDICAL DIAGNOSIS

by John E. Overall and Clyde M. Williams

Veterans Administration Hospital, Perry Point, Maryland, and The University of Florida

DIAGNOSIS is an area of medicine which has been said to invite the use of automation. However, before machines can be programmed to play an important role in medical practice, quantitative models must be selected which realistically represent the processes actually employed by physicians in diagnosing disease, which are feasible from the standpoint of information and computer operations required, and which are accurate and effective as aids to diagnosis. Two general types of models have been discussed recently: (1) a discriminant function model in which the several sign and symptom values are each given an empirically derived weight and a single artificial mathematical measure is obtained as the sum of the separate weighted values, and (2) a frequency-count model in which the relative frequency of occurrence of each possible symptom-disease pattern is considered. Both of these models lean heavily upon concepts developed in statistics and probability theory. The purpose of this article is to discuss these two approaches together with some problems inherent in their applications.

THE DISCRIMINANT FUNCTION MODEL

In an excellent research paper, Crooks, Murray, and Wayne (1959) suggest that the physician, after obtaining a history and eliciting the appropriate physical signs, must decide upon the relative importance that he should attach to *each* diagnostically significant clinical feature of the case. It is perhaps a reasonable analogy to conclude that the process by which the physician makes a clinical diagnosis involves an intuitive application of the principles of discriminant analysis. The physician attaches greater or lesser weight to each clinical feature of the particular case according to his past experience. The weight assigned to each sign and symptom depends upon two types of relationship. Signs and symptoms which correlate highly with the presence-absence of a disease should be given more weight than those which have little relationship to the presence or absence of the disease. However, the weight given a sign or symptom must also depend upon its correlation with other signs and symptoms. For example, two symptoms may each have a very high relationship to a particular disease when considered separately; however, if they are themselves highly related so that the presence of one is almost always associated with the presence of the other, then there is not a great deal to be gained from considering them both, and the weight given the two of them may be a little larger

than the weight which would be given to either one alone. On the other hand, if the two symptoms with high relationship to the disease process were to have little relationship to each other, then the combined weight given to the two of them would be greater than that given either one alone. Thus, the weight assigned to a sign or symptom depends not only upon the magnitude of relationship to the disease process but also upon the magnitude of relationships to other signs and symptoms.

The discriminant function model results in each sign and symptom being assigned an empirically determined weight. The weight assigned to each sign and symptom depends upon its relationship to the disease process and upon its relationship to other signs and symptoms. To the extent that the presence of one sign or symptom can be predicted from knowledge of other sign and symptom values, the diagnostic weight assigned to it is reduced. The discriminant function model yields a diagnostic equation of the type

$$\hat{Y} = aX_1 + bX_2 + cX_3 + \cdots + pX_q$$

in which the coefficients, a, b, c, $\cdots$, are weights proportional to the magnitude of relationship between the particular sign or symptom and the disease process reduced according to the relationship of this sign or symptom to the others represented in the equation, and X_1, X_2, X_3 $\cdots$, are observed sign and symptom values.

FREQUENCY-COUNT MODEL

An alternative general approach has been described in considerable detail by Ledley and Lusted (1959). Instead of assigning a weight to each sign and symptom to yield a single index number, the entire pattern of signs and symptoms is considered simultaneously and the weight or significance of the whole is judged in terms of the probability with which it is associated with each disease. Whereas in the discriminant function approach the aim is to derive artificial mathematical functions which are weighted linear combinations of sign and symptom values, in the frequency-count approach the aim is to determine the relationship of each possible sign-symptom pattern to the diseases in question. The problem of obtaining adequate knowledge upon which to base probability estimates is critical. A large number of instances of *each* symptom-disease pattern must be recorded before adequate probability statements can be made; yet the number of possible *patterns* is much too large to cope with as a single problem. For example, Ledley and Lusted suggest that a single medical specialty may be concerned with at least 400 symptoms and 300 diagnostic categories. This would result in more than 10^{200} possible symptom-disease patterns, without considering the fact that many cases with similar patterns would have to be observed in order to obtain the necessary relative frequency probability estimates. It is obviously unrealistic, as Ledley and Lusted point out, to attempt to determine probabilities for a system in which the number of different categories exceeds the present population of the United States by a factor of 10^{192}. The problem is of a magnitude many times larger than that considered by Shannon when he concluded that there are approximately 10^{120} continuations to be explored in the game of chess and that with this number there would be less than 10^{16} microseconds available for exploring each if the task were to be accomplished in one century (Newell, Shaw, & Simon, 1958). Even admitting *a priori* that only a small fraction of the symptom-disease patterns could occur in nature, it is unlikely that the population could ever supply enough cases or that a computer could be built to handle the data.

SIGN-SYMPTOM HYPERSPACE

Consider a hollow cube of unit dimensions in which 300 tiny balls are suspended. The balls may be conceived to form various configurations, with some spaced very closely in compact clusters and others spaced individually. The location of each tiny ball can be defined in terms of three co-ordinates with the zero point taken as the center of the cube.

Now suppose that the 300 tiny balls represent the mean values for each of 300

diagnostic groups. The three dimensions of the cube may be conceived to represent three independent sign or symptom measures. For example, "up" may represent greater magnitude for one symptom, "back" may represent greater magnitude for a second symptom, and so on. The vertical, horizontal, and lateral position of each disease group is determined by the average values of the three signs or symptoms for that particular disease group.

Alternatively, consider that each of the signs or symptoms represented by the three dimensions of the cube can take only one of two possible values, i.e., up-down, left-right, front-back in the conceptual hyperspace. This corresponds to the case where the symptom is merely recorded as present or absent, without consideration for degree. Let the vertical dimension represent one such dichotomous symptom. If the symptom is always present in a given disease, the tiny ball representing that disease would be placed near the upper limit of the space. If the symptom is never present in the disease, the ball representing the disease would be located at the extreme bottom of the space. If the symptom is present in some cases and absent in others, the location of the marker representing the disease would be determined by the relative frequency with which the symptom is present in the particular disease. In an analogous manner relative frequencies of other symptoms define points along other dimensions of the hyperspace.

Although it is an oversimplification to consider only three symptoms and to conceive of them arbitrarily in right angle relationships in space, the concept of configurations of diseases in multidimensional sign-symptom hyperspace is useful. Statistical procedures are available for constructing from any number of correlated measurements (signs and symptoms) other mathematical measures which are uncorrelated and which have equal variances (see Rao, 1952; Thurstone, 1947). The reasoning applied to the simple three-dimensional case, which is susceptible to physical representation, can be applied with equal validity to the conception of diseases located in hyperspaces defined by many sign and symptom values. The location of each disease in the multidimensional space is determined by the sign and symptom values most characteristic of that disease.

DISEASE CONFIGURATIONS

Conceiving of diseases as arrayed in a multidimensional hyperspace with location determined by average sign and symptom values, we are led to consider the configurations and relationships between diseases. The fact that the diseases are represented in space defined by axes which are uncorrelated within groups does not necessitate the absence of correlation for group means. A high score for a disease on one measure may indicate the likelihood of a high score on another measure. Thus, the group means will tend to form patterns or to fall along lines in the multidimensional space. Given average sign and symptom values for each disease group, it is possible to calculate the distances between the groups in the sign and symptom hyperspace. Certain diagnostic nosologies will be located close together in the sign-symptom hyperspace, while the distance between others will be great. Diseases which are highly similar in effects upon the patient will tend to form discrete clusters. A desirable first step in developing a quantitative classification procedure would be to identify and characterize the major nosologic clusters in the sign and symptom hyperspace. It may well be that the empirically identified clusters will be found to conform to groupings established in current medical practice.

Because the nosologic groups cannot be expected to distribute themselves symmetrically about the center of the sign and symptom hyperspace, it is reasonable to anticipate that the density of points representing mean values for the many diseases will be greater in some directions than in others. For example, in the three-dimensional space described above, some quadrants of the cube may be crowded with markers representing disease groups while other quadrants are sparsely filled. If one were to attempt to pass an arrow through the space in an effort to come as close as

possible to *all* disease groups, it would make considerable difference at which angle the arrow were inserted. An arrow passing through the space in some directions would come much closer, on the average, to all disease groups than an arrow passing through the space in another direction. In fact, there would be one unique direction for which the sum of squared deviations about the arrow would be at a minimum value. Statistical methods are available (Rao, 1952) for developing artificial mathematical variates which have the optimal property of maximizing distances between all groups relative to error variance within groups. This is the multiple discriminant function approach which the present writers believe to have great potential use in medical diagnosis. Through application of this model, an artificial mathematical measure, which is a weighted combination of sign and symptom values, is obtained.

$$Y = aX_1 + bX_2 + cX_3 \cdots pX_q$$

The obtained measure has the optimal property of differing more, on the average, from one disease group to the next than any other linear combination of patient values.

CORRELATED MEASUREMENTS AND REDUCTION IN DIMENSIONABILITY

Absence of correlation is represented geometrically by vectors in a 90-degree relationship. The reason for this is intuitively obvious since lack of correlation indicates that information concerning projection on one dimension will be of no help in estimating projection on the other dimension. Correlation is represented by a smaller angle between two vectors. Two measures which are perfectly correlated are represented by collinear vectors with zero angle between them.

Correlations between signs and symptoms form patterns or clusters. The presence of one particular symptom suggests the presence of several others which are related to it, but provides no information whatsoever about the possible presence or absence of many others. In the conceptual sign-symptom hyperspace, related symptoms form *bundles of vectors* with small angles between them, while vectors representing other signs and symptoms have greater angular disparity from those in the clusters. When two vectors are highly correlated, as are two vectors within a single bundle, one of them will acount for most of the variance in the other. Where several vectors (signs and symptoms) form a tight bundle, a single vector may be used to represent most of the variance in all of them. Methods subsumed under "factor analysis" (Thurstone, 1947) are useful in identifying clusters or bundles of sign and symptom measures and in defining a relatively small number of artificial mathematical variates which can be used to account for a great portion of the variance in the total sign and symptom hyperspace. Perhaps, instead of 400 dimensions representing 400 signs and symptoms, 90 percent of all distances between disease groups can be represented in only 20 or 30 relatively independent dimensions.

THE SEQUENCE OF DIAGNOSTIC DECISIONS

Although it is probably unrealistic to conceive of diagnosis in terms of a single tremendous decision involving consideration of hundreds of symptoms and hundreds of possible categories of classification, it is quite reasonable to conceive of diagnosis as involving a sequence of "small" decisions. If all possible nosologic groups are conceived of as arrayed in a hyperspace of dimensionality equal to the total number of signs and symptoms recognized in medical practice, and if disease classifications tend to *cluster* irregularly in the sign and symptom hyperspace, the first task is to find a relatively small number of sign and symptom measures which will estimate well the vector which maximizes distances between the major nosologic clusters. This procedure is analogous to passing an arrow through the hyperspace in such a way that it comes as close as possible to the centroids of the several major nosologic clusters. The best estimate of the path of the arrow comes from the few signs and symptoms defining vectors which parallel as nearly as possible the direction of the arrow. The first step in developing a quantitative diagnostic proce-

dure might be to identify the major clusters of diagnostic categories, and the second step would then involve identification of a relatively small number of signs and symptoms which will provide maximum distance between the centroids of these major clusters.

At this point, it may be helpful to consider a hypothetical example. Suppose that the degree of presence or absence of 1000 signs and symptoms were recorded for a number of patients in each of 1000 possible diagnostic categories. The mean values for the 1000 diagnostic groups can be conceived to be distributed in 1000 dimensional space. Those diseases which tend to cluster together in the symptom hyperspace will have little distance between them, and those diseases which belong to different clusters will have greater distances between them. Suppose that three major clusters are identifiable: (1) patients with one psychiatric diagnosis are closer in the symptom hyperspace to patients with another psychiatric diagnosis than to patients with nonpsychiatric diagnoses, (2) patients with various infectious diseases are closer to one another in the symptom hyperspace than to patients without infectious diseases, (3) patients with various accident-induced injuries are closer to one another in the symptom hyperspace than to other patients.

Although the three major clusters are initially conceived to be located within 1000 dimensional hyperspace, information on far fewer than 1000 signs and symptoms is necessary to assign a patient with a high degree of accuracy to one of the three groups. Such simple indicators as body temperature, obvious irrationality, and history of recent accident or injury would be useful in such an initial classification.

Once classification into one of the major clusters has been made, the process is repeated, except that this time the aim is to find a reasonable small number of signs and symptoms which will maximize distances between diseases *within* the major cluster. The direction of the vector which maximizes differences between diseases within a major cluster is not likely to be the same as the direction of the vector which was found to maximize distances between the major clusters. For example, "obvious irrationality" may distinguish well between normals and psychotics; however, since most psychotics manifest gross irrationality, this would probably not be a sign which would be most useful in classifying the psychiatric patient into one of the more specific nosologic groupings. Such items as degree of excitement, suddenness of symptom onset, suicidal or homocidal tendencies, and neurologic findings would be more helpful at the second level of classification.

Actually, the complete process of diagnosis may be conceived to include a sequence of several decisions, each of which is accompanied by action which affects the subsequent classification of the patient. For example, Veterans Administration Hospitals are roughly classified as General Medical and Surgical, and Neuropsychiatric. The initial classification of the patient determines the type of hospital to which he is admitted. Within the General Medical and Surgical Hospital, different general types of patients are assigned to different specialists and different wards. Each medical specialty considers many specific diseases and uses many laboratories and special services at one time or another. Since all possible laboratory work cannot be undertaken for each patient, that relatively small number of signs and symptoms which best indicates what further laboratory measures should be obtained is sought by the diagnostician. This amounts to classification of the patient into a subcluster for which certain specific additional information should be obtained. The process of successive classification continues until the patient is finally classified into a group which will profit from treatments T_1, T_2, $\cdots$ T_p. At each stage, the configuration of diseases (within the cluster under consideration) is examined for a dimension along which distances are at a maximum, and this single dimension is defined by a reasonably small number of signs and symptoms.

FEASIBLE STEPS TOWARD QUANTITATIVE DIAGNOSIS

It is one thing to consider applying existing statistical models to restricted problems in medical diagnosis, such as the classification into one of a few diagnostic groups which might be the final step in a long sequence of diagnostic decisions, but it is quite a different thing to consider the use of these models in setting up a comprehensive diagnostic program in which all decisions involved in the diagnostic sequence are considered. It is the purpose of this section to outline some steps which appear feasible in approaching the larger problem of establishing a comprehensive system of diagnostic decisions. The general approach which the writers consider to be practical involves the making of a series of decisions, each decision delimiting further the number of diagnostic classifications to be considered in subsequent decisions. At each stage the amount of information to be considered must be reduced to reasonable magnitude by searching for those few signs and symptoms which provide maximum discrimination between all of the diagnostic groups to which the patient might reasonably belong at that stage in the process of diagnosis. As a result of successive decisions, the number of diagnostic categories is reduced and new measures are sought which maximize discrimination between the remaining groups. The total problem in establishing such a comprehensive system reduces to that of deciding how many decisions are necessary and what information is necessary for each.

The difficulty of establishing such a comprehensive system arises from the magnitude of the total problem. If the system is to be comprehensive in any sense of the word, hundreds of possible diagnoses and hundreds of possible signs and symptoms must be considered. It must start with the fact that the computer is initially faced with a patient who may belong to any one of the hundreds of diagnostic groups and who may have any number of the possible signs and symptoms. The first task in developing such a comprehensive system involves identification of a reasonably small number of patient characteristics which will reduce the total number of probable classifications most effectively. If the problem were smaller, say involving only a dozen or so diagnostic groups and three or four dozen patient characteristics (signs and symptoms), exact statistical procedures would be appropriate. It is our purpose here to indicate some reasonable approximations, not necessarily the only ones.

Consider once more the total sign and symptom hyperspace with the configurations of disease groups determined by their average sign and symptom values. The first step in developing a comprehensive diagnostic system involves the identification of a small number of signs and symptom measures which maximizes discrimination between all of the diagnostic groups. It should not be difficult to determine empirically which single signs and symptoms are of most use in discriminating between the many disease groups. For continuously distributed measures, such as most laboratory data, the ratio of the variance of group means to the total variance of individual patient values (the correlation ratio) is an appropriate measure, and the required computations would not be beyond the capacity of existing data processing units even if the number of patients observed in each diagnostic group was reasonably large. For sign and symptom values which are recorded only as present or absent, an appropriate frequency statistic of the chi-square type could be used at this stage to determine the discriminating value of the separate measures. Only those signs and symptoms which are found to have high potential for discriminating between all possible disease groups should be considered further at this time.

Of those signs and symptoms which are found to have highest discriminating potential, intercorrelations between them may suggest that not all need be used. For example, if two symptoms are found to be highly correlated so that the presence of one almost invariably indicates the presence of the other, then only one of these measures need be included.

One efficient procedure might be to choose the single sign or symptom which was found

to have the greatest power of discrimination between all groups. This measure would be the first candidate to be accepted for use in the initial diagnostic function. The discriminating power of each remaining sign and symptom might then be "corrected" according to the degree of correlation with the first accepted measure. The sign or symptom value which is found to have the greatest "corrected" discriminating power with respect to all disease groups would then be accepted as the second measure to be included in the initial diagnostic function. This process could be continued, "correcting" each measure for correlation with those already accepted, each time choosing the measure with the highest "corrected" discriminating power.

Although choice of measures to be included in the diagnostic function should be governed primarily by empirically determined discriminating power, consideration of clinical practicality cannot be overlooked at any stage in the process. For example, it may be considered impractical to delay the initial diagnostic decision in order to obtain certain pieces of information if a decision of almost as great effectiveness could be reached using other measures which are more readily obtainable. Cost is also a factor with which the physician is ever faced in a practical clinical situation. The advantage acruing from inclusion of a measure at any stage in the diagnostic process must be weighed against the cost.

Turning back to the problem of selecting a parsimonious set of highly discriminating signs and symptoms when these measures are obtained only as present-absent dichotomies, a frequency method analogous to the above step-down procedure appears feasible. The single dichotomous sign or symptom having the greatest discrimination potential could be chosen first. This variable could then be paired with each of the remaining signs and symptoms to identify the one which adds most to the discrimination between all groups. Successive variables could be added, not to exceed four or five in all, to yield a limited number of sign-symptom patterns with high discriminating potential to be used in making the initial diagnostic decision. Use of the frequency model restricts sharply the number of pieces of information which can be considered in a single decision simply because of the tremendous increase in number of sign-symptom permutations accompanying the addition of each new variable. For example, there are 16 possible patterns for four dichotomous measures, while there are 32 possible patterns for five dichotomous measures.

Once the set of measures to be used in making the initial diagnostic decision has been selected, either the multiple discriminant function model or the frequency model can be employed to yield the probability values actually used in the initial classification of a patient. The choice will depend to considerable degree upon the nature of the signs and symptoms to be considered (continuous or dichotomous), upon the number of cases for which complete data is available in each diagnostic group, and upon the number of signs and symptoms to be considered in the initial diagnostic decision. Either model can be used to obtain an estimate of the probability with which a patient in each of the diagnostic groups will have any observed pattern of sign and symptom values. As will be discussed in a later section, the probability that a particular patients belongs to a particular disease group is a joint function of this discriminant probability and the *a priori* probability derived from the relative frequency of occurrence of the several diseases.

As a result of the initial diagnostic decision, the total universe of possible diagnostic classifications is reduced to several subsets. Within each subset, the process of finding the most discriminating sign and symptom values must be repeated to yield the basis for a second diagnostic decision. The method is exactly as described for the first stage except that the number of disease alternatives has been markedly reduced. Although it is possible, it is not likely that the signs and symptoms which were found to be most effective in discriminating between all diagnostic categories will also be most effective in discriminating between the diagnostic categories within a subset.

BASE RATES IN STATISTICAL DIAGNOSIS

Once the discriminating sign and symptom measures have been identified and the relationships between disease group means have been determined, as in the preceding section, the problem of classification of a patient into one of the possible diagnostic groups might appear quite simple. The sign and symptom values for a particular patient could be inserted into the same equations used in determining the locations of the group means. The patient could then be assigned to the disease group from which his sign and symptom value differed least. However, further consideration shows that such a procedure would lead to erroneous classification in many instances. Although 95 per cent of the cases in each diagnostic group may be conceived to fall within a specified radius from the group mean, this does not insure that the absolute number of cases falling outside this radius will be equal for all disease groups. Some diseases occur with much greater frequency than others in the population. Knowing the variances of the discriminant functions, it is possible to calculate the probability with which a *patient who belongs to any specific diagnostic group* will have a sign and symptom value which deviates from the group mean by any specified amount. It is also possible to compute the probability with which *any randomly selected patient* will belong to any specific diagnostic group from the relative frequencies of occurrence of the various diseases in the population under consideration. The joint probability that a randomly selected patient will belong to a specific diagnostic group and will at the same time have a sign and symptom value which deviates by a specified amount from the mean value for that group is given by the *product* of the two independent probabilities.

The importance of automation in medical diagnosis becomes even more apparent when it is recognized that the reference population from which the patient is drawn changes at each stage of the sequence of diagnostic decisions. The probability that a randomly selected patient will have a particular cardiovascular disease differs for the general population and for the population of patients who have been referred to a cardiologist for examination. Each successive decision changes the population to which the patient belongs, and the relative frequencies of occurrence of diseases differ from one population to the next. Effective statistical diagnosis necessitates knowledge of probabilities associated with the incidence of diseases at each of many stages in the diagnostic process. The collection of such data is a formidable, but routine, task which can be successfully accomplished with the aid of electronic data processing facilities currently available. It appears likely that the initial inroad of automation into medical diagnosis will be in the collection and continuous revision of estimates of the relative frequencies of occurrence of diseases in various reference populations. Although the task is routine, the important place of "base rates" in the most sophisticated conceptions of quantitative diagnosis cannot be overlooked.

REFERENCES

Crooks, J., Murray, I. P. C., & Wayne, E. J. Statistical methods applied to the clinical diagnosis of Thyrotoxocosis. *Quart. J. Medicine*, 1959, 28, 211–234.

Holzinger, K. J., & Harman, H. H. *Factor analysis: A synthesis of factorial methods*. Chicago: Univ. of Chicago Press, 1941.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Ed. Psychol.*, 1933, 24, 417–441, 498–520.

Ledley, R. R., & Lusted, L. B. Reasoning foundations of medical diagnosis. *Science*, July 1959, 130, 3366.

Newell, A., Shaw, J. C., & Simon, H. A. Chess-playing programs and the problem of complexity. *IBM J. Res. Develpm.*, 1958, 2, 320–335.

Rao, C. R. *Advanced statistical methods of biometric research*. New York: Wiley, 1952.

Roy, S. N. *Some aspects of multivariate analysis*. New York: Wiley, 1957.

Thurstone, L. L. *Multiple factor analysis*. Chicago: Univ. of Chicago Press, 1947.

(Manuscript received July 18, 1960.)

Many investigators have compared the average performance of individuals and of groups working on a standard task. Of additional interest is the question whether groups perform more effectively than their most competent (or least competent) members would have performed if they had worked alone. Methods of answering such questions have been reported in the literature, but practical difficulties have limited their usage. This paper reports an alternative procedure which will sometimes be appropriate when others are not, and which is based upon the logic of sampling theory as it concerns percentile scores.

A MODEL FOR THE COMPARISON OF INDIVIDUAL AND GROUP PERFORMANCE SCORES

by Ivan D. Steiner and Nageswari Rajaratnam

University of Illinois and University of British Columbia

THIS paper describes a model which can be used to test a variety of null hypotheses concerning the relative performance of individuals and groups.[1]

Early investigations of the problem-solving ability of individuals and groups tested the following null hypothesis: when individuals and members of groups have been randomly selected from the same population, the mean performance of individuals equals the mean performance of groups. Laboratory studies by Watson (1928), Shaw (1932), and Bos (1937) were typical of this approach. The null hypothesis was frequently, though not always, rejected; and the research findings were often interpreted to mean that groups are superior to individuals who work alone because members of groups perform their task co-operatively and have positive, facilitative effects upon one another.

During the past decade it has become increasingly apparent that the null hypothesis which engaged the attention of the early investigators is not the only one which should have been tested, and that the interpretation placed upon their research findings may not always have been completely valid. Taylor (1954), Marquart (1955), Lorge and Solomon (1955), and Faust (1959) have suggested that there are circumstances under which groups could be expected to be more productive than individuals who work alone even if no co-operative or facilitative effects are assumed to occur in groups. For example, when the experimental task is of the "Eureka" type (i.e., tasks for which the correct solution can easily be recognized as being correct), the presence in a group of a single individual who can solve the problem may be sufficient to enable the group to solve it. Under such conditions it may be hypothesized that groups will function at the level of their *most competent members*, rather than at the level of their "average" members.

There is a third possibility which, although it has received very meager attention in the literature, is equally realistic. When successful completion of a task by a group requires unanimous agreement by the members, or requires that each member perform a certain necessary portion of the task, groups in which no co-operative or facilitative actions are assumed to occur should function at the level of their *least competent members*. Furthermore, it is not difficult to imagine situations in which groups whose

[1] The work leading to this publication was done while the senior author was supported by a fellowship granted by The Ford Foundation. The conclusions, opinions, and other statements in this publication are those of the authors, and are not necessarily those of The Ford Foundation.

members operate as do solitary individuals could be expected to perform at the level of their second or third least (or most) competent members. Such might be the case, for example, if the "rules of the game" require a two-thirds majority in order for action to occur, or if group success is contingent upon the proficiency with which some (but not all) of the members can simultaneously move toward a goal. Because tasks can take a wide variety of forms, it is apparent that different investigations of "group effects" may require that different null hypotheses be tested.

A statistical model presented by Taylor (1954) and by Lorge and Solomon (1955) was designed to test the null hypothesis that groups function at the level of their most competent members. However, their model can be used to test other null hypotheses as well. If, in a given population of subjects, the proportion possessing the ability to solve a problem is P, and the proportion not possessing that ability is Q ($Q = 1 - P$), the probability of drawing at random from that population a single subject who will not have the ability to solve the problem is Q. The probability that nobody in a randomly assembled group of size m will have the ability to solve the problem is Q^m, and the probability that at least one member of the group will be able to solve it is $1 - Q^m$. Consequently, it may be said that the most competent members of 100 $(1 - Q^m)$ per cent of all such groups will have the needed ability, and that the most competent members of $100 \times Q^m$ per cent of all such groups will not have it. These percentages are, of course, subject to sampling error when a limited number of groups are under study, but Lorge and Solomon have derived the formula for computing the appropriate error terms, thus permitting a test of the hypothesis that groups are functioning at the level of their most competent members.

For groups like those described above, the probability that less competent members of the groups can solve the problem can also be estimated. For example, $100 \times P^m$ per cent of the least competent members of such groups should have the ability to solve the problem, and 100 $(1 - P^m)$ per cent of the least competent members should be unable to solve it. Through inspection of the terms of the binomial expansion it is possible to determine the percentages of groups in which members at intermediate levels of competence are, and are not, likely to possess the ability to solve the problem. Although the appropriate error terms for these latter cases have not yet been derived, there is little reason to doubt that they can be.

In its expanded form, the Taylor-Lorge-Solomon model would permit m (the number of members in the experimental groups) different null hypotheses to be tested within the framework of a single investigation. But there is a serious limitation to the applicability of the model; it is designed for use only with data of the dichotomous type. If, as is often the case, the quality of individual and group performances are represented by scores on an interval scale, the experimenter can employ the model only if he is willing to sustain the loss of information which is entailed by a conversion to dichotomous scores. What is needed, therefore, is a more general model which will permit the investigator to test the same hypotheses when data are in the form of scores on an interval scale.

A MODEL FOR USE WITH INTERVAL SCALE DATA

Consider a population in which the ability to perform an experimental task is normally distributed. If groups of size m are randomly assembled from this population, and the members of each group are ordered with respect to their ability to perform the task, they will, on the average, divide the population into $m + 1$ equal parts. That is to say, the members of four-person groups will, on the average, fall at the 80th, 60th, 40th, and 20th percentiles of the population (Smith, 1953, p. 572). Consequently, if four-person groups function at the level of their most competent members, they should, on the average, perform better than do 80 per cent of the persons in the population; if they function at the level of their least competent members, their performance could be expected to surpass that of only 20 per cent of the people in the population.

For groups of other sizes, comparable statements can be made.

It is to be noted that the line of reasoning developed above is predicted upon the assumption that an infinite number of groups are drawn from the parent population. In actual practice, researchers seldom investigate more than twenty or thirty groups in a single study. Furthermore, the scores associated with percentiles in the population are rarely known; usually they can only be inferred from data for a sample of persons who have worked individually on the experimental task. For both of these reasons it is necessary to take account of sampling errors which cause discrepancies between sample and population values.

Since only a limited number of groups are studied, the mean of the scores associated with a given percentile in the groups may deviate substantially from the score for the same percentile in the population. But it is known (Cramèr, 1951, p. 369) that, in the limit, the sample values associated with a given percentile are normally distributed around the population value for the same percentile. The standard deviation, σ_P, of this distribution is given by

$$(1) \qquad \sigma_P = \frac{1}{f(\xi)} \sqrt{\frac{P(1-P)}{n}},$$

where $f(x)$ denotes the frequency function in the original population, P the proportion associated with the percentile under consideration, ξ the population value of this percentile, and n the size of the samples. Since we have assumed a normally distributed population,

$$(2) \qquad \sigma_P = \frac{\sigma}{y_P} \sqrt{\frac{P(1-P)}{n}},$$

where σ is the standard deviation of the population distribution and y_P is the ordinate of the unit normal density function at the percentile point associated with P.

Let us suppose that we have group scores, z_g, for a sample of k groups of size m, and individual scores, x_i, for a sample of size N. If the groups are performing at the level of their most competent members, the mean $\bar{z}_g$ of the k group scores is an unbiased estimate of the $(100\ m/m\ +\ 1)$th percentile score for the population. Its standard error, $\sigma_{\bar{z}_g}$, may be estimated from the group scores by the equation

$$(3) \qquad \hat{\sigma}_{\bar{z}_g} = \sqrt{\frac{\sum_g (z_g - \bar{z}_g)^2}{k(k-1)}},$$

Another unbiased estimate of the same population percentile is obtained from the N individual scores by determining the value of x_P, the score at this percentile point in the sample of individuals' scores. The standard error of x_P is estimated by

$$(4) \qquad \hat{\sigma}_{x_P} = \frac{\hat{\sigma}}{y_p} \sqrt{\frac{P(1-P)}{N}},$$

where $\hat{\sigma}$ is estimated from the N individual scores, and P equals $(m/m\ +\ 1)$.

Under the hypothesis that the groups are working at the level of their most competent members, the expected value of $(\bar{z}_g\ -\ x_P)$ is zero, and an estimate of its standard error is given by

$$(5) \qquad \hat{\sigma}_e = \sqrt{\hat{\sigma}^2_{\bar{z}_g} + \hat{\sigma}^2_{x_P}}.$$

A t test may now be employed to test the null hypothesis using the t ratio

$$(6) \qquad t = \frac{\bar{z}_g - x_p}{\hat{\sigma}_e}.$$

The degrees of freedom for this test will equal the number of groups studied plus the number of individuals studied, minus two.[2]

[2] The writers are indebted to Dr. D. Burkholder of the Mathematics Department of the University of Illinois for advice concerning the appropriateness of the t test and concerning the number of degrees of freedom involved when the test is used in the manner described in this paper. It must be noted that sample percentile scores are not normally distributed when the samples are small, and that an assumption of the t test is therefore violated. However, Boneau (1960) has demonstrated that the assumption of normality may be violated to an extreme degree without appreciable effects on alpha values provided ns are approximately equal. When one of the two distributions from which he drew samples was exponential in form, and when sample ns were as small as five, Boneau found that 7.1 per cent of his outcomes fell outside the tabled t value for the 5 per cent level of significance, and 1.9 per cent fell outside the tabled value for the 1 per cent level. When sample ns equaled 15, these percentages fell to 5.1 and 1.4 respec-

Although the above description of the model has been focused upon the case in which group performance is compared with the performance of the most competent members of groups, it is apparent that the same procedure can be employed to compare group performance with that of less competent members of groups.

DISCUSSION

The use of the model to test three different null hypotheses is illustrated by a reanalysis of data reported by McCurdy and Lambert (1952). Eleven individuals and thirteen three-person groups worked on a modified version of the Yerkes multiple-choice apparatus.[3] Six two-position switches were wired so that a light appeared when the total pattern of switch positions corresponded to that on a master control panel. Individuals who worked alone were responsible for all six switches, whereas in the three-person groups each subject was responsible for two switches. Performance was measured by the number of correct switch patterns achieved during a period of 900 seconds. The mean performance of individuals who worked alone was found to be significantly superior to the mean performance of groups ($p < .05$ by an F test).

McCurdy and Lambert suggest that the relative inefficiency of their groups may have been due to the high probability that each group contained at least one member who had been inattentive to the experimental instructions, or who, for other reasons, had performed his portion of the task poorly. Because groups could succeed only if every switch was placed in its proper position, it is not at all difficult to imagine that the progress of groups depended upon the effectiveness of their least competent members. But McCurdy and Lambert's findings cannot be safely interpreted to mean that groups performed at the level of their least competent members, or even that they performed less well than their second most competent members would have done if they had worked alone. A comparison of the mean scores earned by groups and by a sample of individuals permits only intuitive guesses to be made about the performance levels of groups relative to the performance levels which specific members of those groups would have achieved if they had worked individually. The model described above can be employed to obtain more precise information.

Since each of McCurdy and Lambert's experimental groups contained three members, three different null hypotheses can be tested by means of this model. Listed below are the probabilities associated with each of these hypotheses:

Null Hypothesis	t	p
1. Groups performed at the level of their most competent members: $\bar{x} - x_{P=.75} = 0$	4.70	$p < .001$
2. Groups performed at the level of their second most competent members: $\bar{x} - x_{P=.50} = 0$	1.06	$.2 < p < .3$
3. Groups performed at the level of their least competent members: $\bar{x} - x_{P=.25} = 0$	.23	$.8 < p < .9$

It is to be noted that neither the second nor the third of these hypotheses can be rejected at conventional levels of confidence. Even though the mean of the group scores was found by McCurdy and Lambert to be significantly lower than the mean of the scores earned by individuals, the present analysis indicates that there is a substantial possibility that groups functioned as well as their second most competent members would have done if they had worked alone. It is true, of course, that these findings are based upon extremely small samples of individuals and groups; if more cases

tively. It is apparent that use of the procedures outlined in this paper will result in outcomes which are not distributed exactly as is t when samples are small. However, the amount of error involved in using tabled t values is relatively small and can probably be tolerated in most research situations.

[3] McCurdy and Lambert also performed a second experiment which investigated the performance of individuals and groups on the Yerkes apparatus. However, the subjects and the experimental conditions were sufficiently different from those of the first experiment to prevent our combining the data of the two studies. For illustrative purposes we shall deal only with the single study which produced the most significant findings. Although McCurdy and Lambert transformed their productivity scores to achieve greater normality, in the reanalysis of their data which follows we deal with raw scores. The model described in this paper is not designed for use with transformed scores.

had been studied the finding might, or might not, have justified different conclusions. But it is clear that *the available data* do not permit a firm rejection of the hypothesis that groups functioned at the level of their second most competent members.

Taylor (1954) has proposed an alternative method for testing hypotheses of the kind described in this paper. Individuals are randomly assigned to work either individually or in groups. After the experiment is completed, those who have worked alone are randomly arranged into "nominal" groups of the same size as the real groups. To test the null hypothesis that real groups have performed at the level of their most competent members, the mean of the scores earned by real groups is compared (using a t test, for example) with the mean of the scores earned by the highest scoring members of the nominal groups. Although Taylor apparently intended this model as a means of comparing the performance of groups with that of their most competent members, there is no reason why it should not be equally appropriate as a tool for comparing the performance of groups with that of their less competent, or least competent, members.

Taylor's recommendations were followed in a second analysis of the McCurdy and Lambert data. The obtained t values were 2.38, .69, and .17 for the first, second, and third null hypotheses. These t values are slightly smaller, and are based upon fewer degrees of freedom, than are those obtained through the use of the model proposed in this paper. They permit the first null hypothesis to be rejected at the .05 level of confidence; they do not justify a rejection of the second or third null hypotheses. The fact that the Taylor model produced smaller t values than did the Steiner-Rajaratnam model is undoubtedly due in part to the loss of two cases from the sample of individuals who worked alone; only three three-person nominal groups can be assembled from a sample of eleven individuals, and two individuals must, inevitably, be left over.

Although it is believed that the two models should ordinarily produce very similar findings, there are situations in which the model described in this paper may have certain advantages. One such situation has already been encountered above. When only a very small sample of individuals is available from which to construct nominal groups, the loss of even a few subjects as "left-overs" can have an appreciable effect upon obtained findings. This type of attrition in the sample can be especially serious if the real groups are large in size (e.g., if each real group includes eight members, and only fifteen individuals are available for inclusion in nominal groups). Even when there is no such attrition in the sample, use of the Taylor model entails a loss of degrees of freedom which can be avoided by the use of the model described in this paper.

A second advantage of the model described here becomes apparent when one attempts to draw inferences from published research comparing the mean performances of groups with the mean performances of individuals who have worked alone. If, as is usually the case, only the mean and the standard deviation of the individuals' scores are presented, the reader cannot "assemble nominal groups" and test various hypotheses. But if the sample of individuals' scores can be assumed to have been normally distributed, the sample percentile scores associated with different hypotheses can be computed and the model described here can be used to test these hypotheses. Consequently, by means of this model, it is sometimes possible to perform *approximate* tests even when only a summary of the original data is available.

A third advantage of the model concerns its ability to generate hypotheses as well as to test them. When conditions warrant the assumption that group scores depend upon the ability of the ith most competent members of groups, the model permits predictions to be made about the relative performance of groups of different sizes. Thus, if it is the most competent members of groups who determine group scores, three-person groups should, on the average, perform at the level of the 75th percentile of the population from which members have been randomly drawn. Groups of size four, five, and six, drawn from the same normally distributed population, should

perform at the levels of the 80th, 83.3rd, and 85.7th percentiles respectively. By determining the standard score equivalents of these percentiles, one can identify a relationship between group scores and group size which should prevail if group success is determined by the ability of the group's most competent member. (With increasing group size, group scores will increase at a rate proportional to the standard scores which correspond to the indicated population percentiles.) Comparable operations can be performed to derive relationships between group scores and group size on the assumption that group scores depend upon the ability of less competent members.

It is to be noted that no data are needed in order to derive these hypothetical relationships between group size and group productivity; they are logical consequences of the model itself, and can be deduced even before an experiment is begun. Consequently, the model can be employed to infer relationships in addition to those which it is explicitly designed to test.

REFERENCES

Boneau, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, 57, 49–64.

Bos, Maria C. Experimental study of productive collaboration. *Acta Psychologica*, 1937, 3, 315–426.

Cramèr, H. *Mathematical methods of statistics.* Princeton: Princeton University Press, 1951.

Faust, W. L. Group versus individual problem-solving. *J. abnorm. soc. Psychol.*, 1959, 59, 68–72.

Lorge, I., & Solomon, H. Two models of group behavior in the solution of Eureka-type problems. *Psychometrika*, 1955, 20, 139–148.

Marquart, Dorothy I. Group problem solving. *J. soc. Psychol.*, 1955, 41, 103–113.

McCurdy, H. G., & Lambert, W. E. The efficiency of small human groups in the solution of problems requiring genuine co-operation. *J. Pers.*, 1952, 20, 478–494.

Shaw, Marjorie E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, 44, 491–504.

Smith, K. Distribution-free statistical methods and the concept of power efficiency. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: The Dryden Press, 1953.

Taylor, D. W. Problem solving by groups. In *Proceedings of the XIV International Congress of Psychology, 1954.* Amsterdam: North Holland Publishing, 1954.

Watson, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, 23, 328–336.

(Manuscript received May 25, 1960.)

". . . this very memory itself is mind (for when we give a thing in charge, to be kept in memory, we say, 'See that you keep it in mind'; and when we forget, we say, 'It did not come to my mind,' and 'It slipped my mind,' calling the memory itself the mind); this being so, how is it that when with joy I remember my past sorrow, the mind hath joy, the memory hath sorrow; the mind upon the joyfulness which is in it, is joyful, yet the memory upon the sadness which is in it, is not sad? Does the memory perchance not belong to the mind? Who will say so? The memory then is, as it were, the belly of the mind, and joy and sadness, like sweet and bitter food; which, when committed to the memory, are as it were, passed into the belly, where they may be stowed, but cannot taste. Ridiculous it is to imagine these to be alike; and yet are they not utterly unlike. . . ."

ST. AUGUSTINE, *The Confessions of St. Augustine*

CRITIQUE AND COMMENT

An initial concern with the problem of the one and the many has led to the Leviathan studies, a series of investigations into internal dynamic governing and deciding processes of large social groups, in particular of complex organizations. In scope, these studies aptly fit the subtitle of their namesake, Thomas Hobbes great work *Leviathan* or *The Matter, Form, and Power of a Commonwealth.* In his introduction, Hobbes said, "Art goes yet further, imitating that rational and most excellent work of nature, *man.* For by art is created that great *Leviathan* called a *Commonwealth,* or *State,* in Latin *Civitas,* which is but an artificial man; though of greater stature and strength than the natural, for whose protection and defense it was intended; and in which the sovereignty is an artificial soul, as giving life and motion to the whole body." Here is an interim report of these studies that are artificially creating the workings of a complex organization by simulating social processes on the computer.

THE LEVIATHAN TECHNIQUE FOR LARGE-GROUP ANALYSIS

*by Sydney C. Rome and Beatrice K. Rome**

System Development Corporation, Santa Monica, California

THE Leviathan studies are a series of investigations into internal dynamic governing and deciding processes of large social groups, in particular of organizations that possess complex structure and perform serious productive work. Examples of such groups are a large military command, a governmental agency like the U. S. Internal Revenue Bureau, or an industrial combination like an international oil corporation.

On the whole, serious studies of social process have been confined to small or primary groups in which each member communicates often with every other and in which the activities of a group can be conducted face to face. Scientific study of large-group processes has been rare. Moreover, in the literature of small-group dynamics, the larger number of investigations has been concerned with interactions between individuals; for example, how one individual influences another or how he perceives other opinions of himself. In diagraming the subject matter of such investigations, we might show an individual situated in a crowd. Concern is primarily with activities proceeding in time, that is, genetic activities and the formal or morphological structure of groups is ordinarily not given much attention.

How a group is organized can, however, obviously make a difference to how it operates. Thus, if 19 people assemble in a classroom, they may organize themselves as a class of 18 and an instructor, while the same people on a playing field might group themselves into two teams and an umpire. Social studies for which formal structure is of the essence include investigations of status and role and of communication nets.

Investigations of small-group dynamics, whether or not morphological structure is emphasized, are generally restricted to systems with only a single level of hierarchy. Usually persons are the individuals, and the social structure comprises an over-all pattern of interactions among them. The

* Presented at the International Congress for Logic, Methodology and Philosophy of Science, Stanford, California, August 24 to September 2, 1960.

social structure has properties which do not characterize the individuals.

Large groups are functionally and structurally different from face-to-face groups. Three differences may be noticed. (1) A large group can contain many individuals whose contributions to the group are not distinctively dependent on particular personal differences. Just who delivers a written message or fills a carton often makes no difference to a large producing group, provided only that such tasks as these are indeed accomplished as required. (2) Secondly, it is physically not possible for every member of a large group to enter into two-way relations with every other member within any reasonable period of time, not even with the aid of intercommunicating equipment. Much of the vital activity of large groups proceeds through one-way communication and through intermediaries. (3) In the third place, large producing groups exhibit structures in which entire face-to-face groups themselves constitute autonomous individuals. Thus a command council, its members acting face to face, can direct the work of other groups that function face to face.

These characteristic modes of functioning and of structuring indicate that *how* large military, governmental, and industrial organizations are organized can vitally affect their dynamic processes. One reason why formal organization affects dynamic process lies in the relationship between governing and productive processes. Consider the respective time flows. Productive processes advance like the progress of a wave front. Simultaneously, events in productive processes can occasion demands for control decisions. Once such events are sensed, their effects can rise into the governing structure, each kind with its characteristic rate of flow. Decisions are made, commands are issued, and interventions in the productive processes follow. Such temporal movements upward and downward in the governing process resemble, not the progress of a wave front, but far more the oscillation of particles within a wave. Frequently, the productive work that occasions an activity of control will itself have been completed before the governing decisions will have been brought to bear on later generations of productive work.

It would be oversimple to suppose that governing and technological processes are just two self-contained systems. A productive system can, of course, for some purposes, be viewed as a simple combination of subsystems, the output of one feeding into the input of its successor, while the aggregate of subsystems comprises a simple net like a nerve net. But a governing system, in contrast, rarely exhibits such aggregational simplicity. Myriad activities temporally proceed in parallel at diverse levels of organization. Novel synthesis of intelligence, of strategies, of norms, of responsibilities, and of spans of authority emerge and hold sway over and over again at successively higher levels of integration. Each higher-level synthesis absorbs lower-level integrations, in analogy to the organization of molecules in cells, cells in organs, organs in animals.

We have argued elsewhere (Rome & Rome, 1960) that neither classical symbolic logic nor existing mathematical formulations of social processes are adequate at present to provide analytic predictive control over such multilevel processes. Accordingly, we are developing a new technique for large-group studies. This technique combines two facets, first, a hypothetico-deductive formulation in terms of general rules, and second, a logical, analogue simulation on a large digital computing machine. On the computer we can obtain precise systemic consequences of various constellations of general social rules. In the remainder of the present paper, we shall describe broad aspects of our computer formulations. Approximately half of the programs are now checked out and operating on the IBM 709 computer, and most of the remainder has been formulated.

COMPUTER FORMULATIONS

Ordinarily, if one wishes to model social process on a computer, he develops families of formulae which are then probed in unison by Monte Carlo sampling techniques. With these cybernetic or dynamic programming

can be used to provide optimal adaptation to social objectives. However, such techniques are not suitable for use in the disinterested study of large organizations for two reasons: they are inadequate to model the multilayered hierarchical structures we are describing; secondly, in the general theory of organizations, univocal optimization is largely irrelevant.

For these and other reasons, we have decided to employ a kind of isomorphic or analogue modeling. We first invent an arbitrary flow of productive work. Governing processes are viewed as capable of modulating this primary productive flow. The arbitrary flow of work at the base of the entire simulation enables us to capture something of the brute—the surd character of actual productive social endeavor.

Next we introduce a very limited set of individual personal characteristics to represent persons (workers, government employees, enlisted men) who do directly productive work. These include skill profiles, differential contributions in performing work, and variations in making productive contributions.

At this stage, we then introduce what an industrial or military sociologist would call formal organization. We can simulate most varieties of abstract formal structure, with more than five echelons of command, and associate such structure with simulated producers and the arbitrary work process in many different ways. For example, we can model an organization structured according to stages of completion of work, starting with a division of the organization for initial processing of raw materials, one for intermediate processes, and a third for final processes. Or again, we can model according to process flow, with each subprocess assigned to a specific first-level officer or supervisor. Thirdly, we can model according to skills, after analogy to departments of engineering, mathematics, economics, and production. Fourthly, randomly generated organization can be modeled just as readily. Any of these four kinds of organizational structure, as well as others, can be made to develop or change according to general criteria that we can impose.

As information flows through the formally acknowledged hierarchy, raw data are subjected to assessment by criteria supplied from superordinate levels of decision and control. Hence, over and over, at many nodes in the governing pyramid, data are transformed. Ever more general statements are generated about more particular statements. Between the more general and the more particular statements there can be modelled in the computer a radical inhomogeneity, or difference in kind—a discontinuity. In consequence, when intelligence is filtered at higher and higher levels, initial raw data are subjected to successive radical resynthesis. In this way we model the divergency of concerns, interests, and objectives that occur among diverse echelons of responsibility in large productive social groups. Since the criteria can change according to general rules, we can also model the development of norms.

Each node in the formal executive structure is occupied by an individual. This individual is modelled with many more particular characteristics than are given the producers. He also has a signature at least 16 bits long that distinguishes him from all other individuals. As needs for strategy and decision arise that cannot be satisfied promptly by the formal organization, an individual decision maker in the pyramid of command seeks another that most nearly fulfills his particular conception of a suitable associate. The first individual then makes a request of the other. As these requests succeed or fail, a web of informal associations becomes established in the computer. An example will make clearer what this means. If decision processes are slower in the formal organization than in the informal, after a while major decisions will tend to be made by informal coalitions. If now executive replacements are made in the formal organization on the basis of informal acquaintance among the officers, the informal organization may eventually take over the formal organization. But this is not all. Many of the out group will be candidates for reconstituting the informal net, for they will form their own in-groups. If formal decision making continues to be

slower than informal, the displaced group may in large part eventually complete a cycle by returning to formal power via informal associations. The important point is not merely that we can model this, but that we can explore precise experimental controls for inducing this over-all effect and for enabling us to determine experimentally many of its qualitative and quantitative characteristics.

A large social organization can be viewed as an elaborate set of games being played out in parallel. The respective individuals and their coalitions have demands that, according to conditions in the organization, are active, latent, or unavailable, and are evoked differently under different conditions. These are satisfied to extents that are partly determined by the degree to which the organization can maintain itself above a minimum level of economic survival. The organization is constantly making side payments to the individuals and to the coalitions. These payments—monetary, status and policy payments—are themselves modulations of decision and productive processes.

At each decision node, various kinds of pressure are exerted and resisted in many ways. Thus the organization can also be viewed as a moving equilibrium of decision processes. In consequence, at each decision node, various kinds of demands, contributions, and evaluations issuing from subordinates, staff, associates, superordinates, and from private internal demands are continually being synthesized. These syntheses result in developing strategies, successive decisions, and consequent demands and commands.

The entire complex of these decision processes at individual nodes, each with its characteristic rate of flow and possible alternative complexities, reaches continuing synthesis as the computer enacts out the history of a specific organization that is being investigated.

Sometime in 1961, we also expect to have a facility for introducing perhaps eight or more live decision makers who will actively intervene with the computer in real time in the roles of intermediate executives or of top management, while our experiments are performed.

It should be observed that we can repeat our experimental runs, duplicating all imposed conditions, but modifying or eliminating some basic module of the governing process. For example, we can operate our simulation with and without informal organization, or with widely differing time lags in any given decision factor, or with different traffic capacities or other modes of operation at the several levels of command. In this way, by subtracting a basic module or by treating it as autonomous and modifying only this module, we shall be able to factor out system effects of basic modules in the governing process.

Ordinarily, one currently views a communication net in the manner associated with the well-known experiments of Alex Bavelas and others. In a large organization, one would want to explore more complex nets. Different links in a sophisticated net are used for different purposes, present diverse obstacles, promise diverse rewards, subtly overlap one another. Then, too, command is not always as effective as an executive wishes it to be. "Inertias" in the productive system can poison essays at control; productive systems are highly selective in their sensitivity to effective control.

A sophisticated network of nodes and links that overarches a technological system suggests two basic kinds of investigations of complex hierarchical governing structures. We can focus on the complex links between the nodes, studying the effects of diverse manners of operation through the links and organization of the links, especially the effects of differential time lags and of radically distinct kinds of messages. But equally we can place individuals with different characteristics at the particular nodes. For example, we can study the effects of progressively less ability to delegate upon the part of decision makers as they occupy more exalted positions in the command pyramid, or of increasingly trivial choices of subjects for making decisions.

Focusing at will on the links between nodes or on the characteristics of individuals at the nodes in our decision pyramid, we

shall explore principally the effects of social pathology on time flows. What are the consequences of bad, irrelevant, or mistimed decisions? What differences do episodic or cyclical occurrences of these make? What will be sought and measured in our experiments will be time rates, cycling effects, and, if we are sufficiently artful, perhaps phase effects resulting from precisely specifiable forms of social degeneration.

REFERENCE

Rome, S. C., & Rome, Beatrice K. Formal representation of intentionally structured systems. In A. Kent (Ed.), *Information retrieval and machine translation*. Proceedings of the International Conference for Standards on a Common Language for Machine Searching and Translation. New York: Interscience Publishers, 1960, Ch. 12, Part I.

(Manuscript received November 15, 1960.)

"Reflecting on my experience, I find myself agreeing with the eminent Cambridge philosopher, Dr. C. D. Broad, 'that we should do well to consider much more seriously than we have hitherto been inclined to do the type of theory which Bergson put forward in connection with memory and sense perception. The suggestion is that the function of the brain and nervous system and sense organs is in the main *eliminative* and not productive. Each person is at each moment capable of remembering all that has ever happened to him and of perceiving everything that is happening everywhere in the universe. The function of the brain and nervous system is to protect us from being overwhelmed and confused by this mass of largely useless and irrelevant knowledge, by shutting out most of what we should otherwise perceive or remember at any moment, and leaving only that very small and special selection which is likely to be practically useful.' According to such a theory, each one of us is potentially Mind at Large. But in so far as we are animals, our business is at all costs to survive. To make biological survival possible, Mind at Large has to be funneled through the reducing valve of the brain and nervous system. What comes out at the other end is a measly trickle of the kind of consciousness which will help us to stay alive on the surface of this particular planet."

ALDOUS HUXLEY, *The Doors of Perception*

A quarter of a century ago, Benjamin Lee Whorf brought forward his challenging hypothesis that the structure of the language in a given culture determines the "world view" of the culture. This thesis has been vigorously debated by anthropologists, psychologists, and linguists, but not much has been done to indicate how such an issue could be decided. The possible implications of Whorf's hypothesis are spelled out in this paper in an attempt to sharpen and differentiate them. In particular, the author distinguishes between the descriptive and explanatory components, and also between the cognitive and ethical aspects of a world view.

SOME REMARKS ON THE WHORFIAN HYPOTHESIS

by Gerard A. Radnitzky

University of Gothenburg

IN his lucid and penetrating article "A Systematization of the Whorfian Hypothesis," Joshua Fishman (1960) points out that the problems connected with the so-called Whorfian hypothesis call for an *interdisciplinary* approach by linguists, psychologists, sociologists, etc. This is certainly so. The "hypothesis" is also of obvious interest to students of philosophy and perhaps they too could make a contribution. In this paper it is suggested—in continuation to Fishman's suggestion (1960, p. 336) that the impact of language *per se* ought to be more fundamental than that of some language features—that the fundamental units ought to be "conceptual systems," which qualify as special languages. It is suggested that considering such systems from the viewpoint of the philosophy of science may provide a useful point of departure for the conceptualization of problems connected with Whorfian effects, and some tentative general proposals are made.

1. THE METASCIENTIFIC FRAME OF REFERENCE

The apology for beginning by sketching, in *nontechnical* terms, an approach of the philosophy of science is that, while science is but a continuation of common sense inquiry, the various "conceptual systems" are obviously clearest in scientific contexts; and hence as it is easier to discuss such systems than any others, they may be used as paradigms.

1.1 The results of scientific research may be roughly classified into two sorts: a body or system of knowledge, and a set of conceptual systems which are used as instruments for describing and explaining (e.g., a botanical taxonomy or Newton's "theory," etc.). The continuous interaction between the two is of the greatest importance in the development of a science; the conceptual systems are necessary for building up the body of knowledge, and a certain body of information is a prerequisite for constructing the systems.

1.2 It is imperative to distinguish clearly between descriptive system and description. (Explanatory systems or "theories" and explanations are treated by implication.) A descriptive system is an instrument for producing a systematic description. For example, a map is a description of a certain territory; its legend epitomizes the descriptive system used. The relation between the two is a special case of that between language and discourse (speech). A descriptive system, as a language with syntactical and semantical aspects, "contains" sets of concepts (in the logical sense) and sets of rules stipulating how descriptive discourses are to be produced.[1]

[1] Concepts in the logical sense are here distinguished both from attributes and from concepts in the psychological sense. Some paradigm sentences showing how the terms are used will help to suggest the idea. 'The concept (logical) of world-line was introduced by Minkowski,' 'After the time of Lavosier chemists no longer used the

Let us shortly consider an example of the development of a descriptive system. A physical "geometry" is an instrument for producing systematic descriptions of physical objects in the universe. Before Poincaré, no distinction was made between geometrical description and the descriptive system used for it. When it became apparent that Euclidian geometric systems can be adequate only with respect to small spaces, new descriptive systems with greater realm of application were constructed. In addition, the postrelativistic geometric descriptive systems became an integrated part of a combined descriptive-explanatory system. Other examples would be the various kinematics, botanical taxonomies, the periodical system in chemistry, psychological typologies and dimensional systems, etc.

The special characteristic of such descriptive or explanatory systems that interests us here is that it is irrelevant in which natural language they are "presented" and the description or explanation formulated, i.e., the discourses in question can be translated from English into Japanese (say) without anything essential being lost. Although their realm of application is limited, the systems belong to what might be called (with a somewhat highflown title) the universal language of science. They are discarded, changed, or reconstructed in consequence of considerations of their adequacy (Törnebohm, 1957; 1958).

1.3 What is being described depends on the inquiry at hand (and often such descriptions are being preferred that can be explained). *How* something is being described depends on the choice of one descriptive system out of those available. And the preference for one such system has to be justified by reference to considerations of the adequacy of the descriptions achieved, i.e., mainly their "correctness" and their "relevance."

1.4 The descriptive (or explanatory) systems may also be made the subject of a pragmatic study, i.e., the emphasis is placed upon the communicants and their attributes instead of treating them as schematized concepts or abstracting from them altogether, as has been done above. We may, for instance, be interested in the relation between descriptive systems and sets of language users or speech communities. In scientific contexts, we find the unit of grouping to be more or less esoteric groups (e.g., in connection with "laboratory discourse") with a single individual who functions both as communicator and communicatee as a special case (soliloquy common in the context of research). Systems may become "built into" a sublanguage in the sense that their use is well-established and common in the speech community, and the discourses "presenting" them and "giving" the corresponding descriptions or explanations can be formulated in the sublanguage with relative ease and felicity (Fishman, 1960, p. 335). They may be "built into" some natural language, in which case the speech community becomes identical with the set of "normal" or "representative" language users of that language. And they may be idiosyncratic to one natural language or to a group of such. For example, the system for describing (phenomenal) colors with respect to hue, that is based on the color pyramid, may very well be "built into" Western languages but not into some others (Fishman, 1960, p. 330). But it may of course be possible for a *S* of another speech community or language group to become acquainted with and learn to use such a descriptive system without learning a whole sublanguage or natural language.

---

concept of Phlogiston; yet it did not cease to exist, it was used in texts on the history of science.' To find out whether a certain concept (logical) occurs in a text we have to ascertain whether the text contains a term to designate it; while to find out whether an object has a certain attribute, we have to (say) look at it. To find out whether *S*s "possess" a concept (in the psychological sense) of (say) "blue," we observe, e.g., their sorting behavior, etc.; and we may propose a psychological explanation in which a *logical* concept of the psychological concept of "blue" is utilized. If a *S* has knowledge of a descriptive system, he has at his disposal also certain psychological concepts corresponding to those of the system. To "have" a concept in the psychological sense it is not necessary that its expressor be vocalized and there may probably be degrees in the development of concepts from being identified through vague recognition to having a well-articulated label. (On psychological concepts see Adams, 1953.)

2. IMPLICATIONS OF THE METASCIENTIFIC CONSIDERATIONS FOR THE STUDY OF WHORFIAN EFFECTS IN "COGNITIVE" BEHAVIOR

2.1 Describing

'Describing' or 'describing behavior' is here used in the widest sense, including not only the producing of discourses but also manipulation, e.g., sorting behavior, and even perceiving something as an object of a particular *kind*. It is a truism that how we "describe" depends on the choice of the descriptive system (however rudimentary it may be and however unaware the person may be of the fact that he makes a choice), and hence partly on what descriptive systems are readily available.

In the context of scientific inquiry, the choice or construction of a descriptive system is guided by considerations of the "adequacy" of the descriptive discourse achieved by using it. The discourses making explicit these considerations belong to the philosophy of science (i.e., they talk *about* scientific discourses and hence stand in *meta* relation to them), and are independent of the natural language used in the above-mentioned sense of being translatable. A study of their pragmatic aspects belongs to the psychology of scientific endeavors, but is not likely to be relevant for the study of Whorfian effects.

Outside scientific context, the choice of the system will follow habitual patterns. By ascertaining how various sorts of empirical phenomena are "commonly" described in a speech community, we get the material necessary for ascertaining what descriptive systems are "built into" the sublanguage (or language). A description (classificatory, typological, etc.) of a sublanguage with respect to the descriptive systems "built into" it is "fruitful" if it makes possible the formulation of many empirical laws (descriptions of observed regularities among empirical phenomena) correlating sublanguage with describing behavior. The next task is to explain these laws, i.e., to construct a suitable explanatory system which among other things explains how such facilitation comes about. It is certainly advantageous to separate the task of describing and that of explaining; and in the subtask of unearthing and reconstructing the descriptive systems and classifying them in order to arrive at a description of the sublanguage, philosophers of science as well as empirical semanticists might prove useful assistants to social psychologists.

2.2 Explaining

'Explaining' is here used so as to exclude the so-called subsumptive explanations, in which a description is derived via a deductive or inductive system from more general descriptions since these trivial cases will be covered by the foregoing; but it includes the producing of discourses "giving" explanation sketches, however rudimentary, and even the process when someone "thinks two phenomena to be causally related." (Prediction is being treated by implication since it differs only in its pragmatic aspects from explanation [postdiction].)

There will be a close parallel to what has been said concerning "describing." In the context of scientific inquiry, choice (or construction) of an explanatory system is guided by considerations of the "adequacy" of the system, e.g., "systematic power," requirement that the descriptive sentences "derived" by means of it are confirmed to a certain degree, that it "connects" confirmed laws, simplicity, etc. (Törnebohm, 1957, p. 136 ff). Again discourses discussing the adequacy of alternative systems explaining the same description of some observed regularity among empirical phenomena belong to the philosophy of science and are not only in principle but technically translatable into any natural language sufficiently rich. Nor is a study of the pragmatic aspects of the process of explaining likely to involve considerations of Whorfian effects.

In connection with the construction of explanatory systems, the "philosophical frame of reference" attracts attention. By 'philosophical frame of reference' is here meant the general approach adopted to "causal analyses," i.e., the answers to the questions: what common properties have the "principles" and what common features the schematizations and model-constructions of a set of explanatory systems, e.g., those con-

structed or used by a certain scientist, and what character he ascribes to them. Even these "metaphysical presuppositions," however, are discussed in metadiscourses belonging to the philosophy of science and independent of natural language in the above-mentioned sense of translatability. However, unless considerations of adequacy and the justification of the adequacy criteria employed is made explicit, there *may* be uncontrolled influences from the explanatory systems customary in the speech community (whether everyday or "laboratory"), and the causal terminology commonly used may "suggest" a special sort of explanatory system.

Outside the context of scientific inquiry, the study of the way in which certain phenomena are "typically" explained in a speech community will again provide the material for ascertaining what explanatory systems can with ease or felicity be formulated by means of its linguistic resources. The study of Whorfian effects in this field is a *part* of the general pragmatic question of what factors both in the semantic environment and the personality of a language user correlate with the explanatory systems he tends to use in everyday situations and the uncontrolled "metaphysical presuppositions" he tends to maintain. Again the formulation of specific hypotheses will have to be preceded by an unravelling and reconstructing of the systems and presuppositions involved.

The requirement set up by historians of philosophy, particularly of those periods when philosophy still encompassed science, to have extensive knowledge of the commonly employed "conceptual frames" of the period concerned goes in the same direction.[2] Unless we have unearthed such systems we have no chance of giving an adequate rendering of the philosophical or protoscientific tenets of these periods.

This leads us to make, in passing, some remarks concerning systems of logic. A good many philosophers have been intrigued by the question whether the syntactic structure of natural language influences philosophical doctrines. (See, for example, Nietzsche's dictum ". . . by the grammatical structure of a group of languages everything runs smoothly for one kind of philosophical system, whereas the way is as it were barred for certain other possibilities" sympathetically quoted by Flew [1953, p. 3]). Or that Aristotelian syllogistic logic "reflects" aspects of Greek language structure. (In a sense the structure of the metalanguage used in connection with a proof is relevant; for instance when '*p* or not-*p*' is proved in an axiom system and a natural language is used as metalanguage, which so to speak already postulates what has to be proved. However this is not what is intended by the philosophers in question.) An axiom system, a "system language"—in contradistinction to a natural language—*constructed* by mathematicians, theoretical semanticists, etc., is strictly independent of any natural language. With a suitable (logical) interpretation, certain systems serve as instruments for reconstructing a discourse with respect to its deductive aspects. A particular system so interpreted may or may not make possible an adequate reconstruction of that aspect; however, it is difficult to see what risk there can be of "linguistic relativity" for a system language.

3. WHORFIAN EFFECTS IN "NONCOGNITIVE" BEHAVIOR

By 'noncognitive behavior' we wish to refer to the producing of discourses with *i.a.* emotive (expressive or evocatively emotive) function, including valuational, moral, etc., discourses, and approving-disapproving behavior in the widest sense. A hierarchy of values, whether organized into a value system or not, must be capable of formulation at least in part if it is at all to be articulated and "consciously" adopted. A value system characteristic or typical of some speech community may, like the descriptive and explanatory systems, become "built into" some sublanguage in the sense that value judgments in harmony with it can be formulated in it with comparative ease and felicity. It is to the interest of the speech community to preserve attitudes in this way, and learning the sublanguage (and hence becoming a member of the speech com-

[2] See Enriques (1936, p. 45 ff); Enriques and Santillana (1936, p. 95 ff); and Copi (1958, p. 45 ff).

munity) may at the same time be a socializing to the value system in question. It is a truism that it may be difficult, verging on the impossible, to formulate (in a nonsatirical manner) a value judgment whose expression is facilitated in one sublanguage, in another whose "built-in" value system is discordant with that of the former; for example, to talk about a criminal act approvingly not in the language of the gang but in the language of the church.

It appears plausible that in making value judgments we often first make a "verdict" in the form of a vaguely felt approval-disapproval, and that, provided we have two language media at our disposal which differ with respect to the ease and felicity with which this "feeling" can be articulated in the form of a value judgment (whether by an utterance or by subvocal soliloquy), we choose that medium that facilitates the formulation and whose "built-in" value system is congenial with it.[3]

Such considerations suggest that evaluating behavior in general may exhibit rather pointed Whorfian effects. Again, a preliminary to the framing of specific hypotheses would be the unearthing and systematizing of the hierarchy of values typically adopted in a certain speech community. (The interlinguistic design of study will here too be only a special case.) This variant of the "Whorfian hypothesis" does not appear to have received very much attention so far from social psychologists. Yet such studies would be of great interest to students of moral philosophy and axiology in general.

[3] Recognition of the facilitative connection between sublanguage or language and value system, including moral standards, is sometimes forthcoming in the psychological and in the philosophical literature. (Lagarde-Quost, 1947, 1948; Hare, 1952, p. 121)

REFERENCES

Adams, J. Concepts as operators. *Psych. Rev.*, 1953, 60, 241–251.

Copi, I. The growth of concepts. In P. Henle (Ed.), *Language, thought, and culture.* Ann Arbor: Univ. of Mich. Press, 1958.

Enriques, F. *Il Significato della Storia del Pensiero Scientifico.* Bologna: Zanichelli, 1936.

Enriques, F., & de Santillana, G. *Compendio di Storia del Pensiero Scientifico.* Bologna: Zanichelli, 1936 (1953).

Fishman, J. A systematization of the Whorfian hypothesis. *Behav. Sci.*, 1960, 5, 323–339.

Flew, A. (Ed.) *Logic and language (second series).* Oxford: Blackwell, 1953.

Hare, R. *The language of morals.* Oxford: Clarendon Press, 1952.

Lagarde-Quost, P. The bilingual citizen. *Britain Today*, 1947, 9, 13–17; 1948, 10, 15–19.

Törnebohm, H. On explanations, predictions, and theories in Physics: a case study. In H. Törnebohm, *Fysik och Filosofi.* Acta Universitatis Gothoburgensis, vol. LXIII, 1957.

Törnebohm, H. The space-time theory within the special theory of relativity. An essay in reconstructive analysis. *Methodos*, 1958, 10, 1–24.

(Manuscript received November 28, 1960.)

History is simply a piece of paper covered with print; the main thing is still to make history, not to write it.

BISMARCK

Anybody can make history; only a great man can write it.

OSCAR WILDE

BOOK REVIEWS

Ralph W. Stacy. *Biological and medical electronics.* New York: McGraw-Hill, 1960.

As Professor Stacy states in the preface, his book is a textbook. The book's contents are taken from a course given in biophysics at Ohio State University. It is a textbook for experimentalists and clinicians in the biological sciences, introducing them to the area of instrumentation and the problems therein. Professor Stacy fulfills the goal he has outlined in the preface.

The two best features of the book are the definitions of the terminology used in instrumentation and the bibliography. Dr. Stacy's uncomplicated style indicates that he is well aware of the background of the readership. The chapters on the theory of measurements and fundamentals of electricity are lucid and are essential for an introductory text.

The section on vacuum tubes could have been done better but I feel that the author felt inhibited realizing his audience would not be sophisticated in terms of electronics.

A shortcoming of the chapter on electronic circuitry is the scarcity of "handy circuits" which the author has found valuable in his own work or has run across from other sources. Handy circuits are circuits which solve some of the problems in experimentation. Examples of these circuits are hash filters, time delay circuits, and preset counting circuits, to mention just a few.

The chapters on detecting and sensing elements and recording devices give a good orientation in these areas. But the two most interesting chapters are those on trouble shooting instruments and computers as laboratory instruments. Trouble shooting even for the best of us sometimes is a frustrating experience. Knowing what the difficulty is and being able to find out where and what is causing the difficulty is an adventure in deductive reasoning. Trouble shooting also demonstrates to the observer whether a measurement (regardless of the device used) is meaningful or not. From the reviewer's experience, it has been interesting to watch people make measurements and observations that have little or no bearing on the difficulty. The person facing a trouble shooting task should know (1) how the apparatus should work and (2) exactly what is wrong. Simply to say "transients!" is frightening but also of little value.

I hope that as the book is revised the chapter on computers is kept up to date. The incorporation of computers in biological work is exciting and holds great promise. Already in basic research they have enabled the experimenter to gain insight into neurophysiological phenomena. In the area of clinical medicine, they have established themselves as useful diagnostic tools.

Professor Stacy's book, in conjunction with Whitfield's *Electronics for Physiological Workers*, provides the beginning of a good library in electronics and instrumentation for students and workers in the biological sciences.

Bertram Peretz
Mental Health Research Institute
The University of Michigan

Christian B. Anfinsen. *The molecular basis of evolution.* New York: John Wiley & Sons, Inc., 1959.

Although supporting evidence was meager, it had been assumed for many years that evolution must ultimately rest on a chemical basis in which the experience of the species is written and is transmitted in a stable form from generation to generation. While the nature of the genetic material remained unknown, it became apparent that the individual acted as a carrier of these blueprints more durable than himself and that in the solution to this problem was the key to replication itself. Over the past 50 years, many biological models have been proposed to explain genetic phenomena, but it has been in the last 10 years that our understanding at the molecular level has become clarified to some extent. The research area

of what has been called chemical genetics has become an exciting and rapidly advancing one. In a short period of time a new language has arisen which is foreign to classical chemist and classical geneticist alike. *The Molecular Basis of Evolution* is welcome in that it represents a timely source book of the current status of this field and also serves as a lexicon for the unintiated.

The reader will find in roughly historical sequence a brief review of Mendelian genetics, the implication of the cell nucleus, then the chromosomes, the correlation of Drosophila phenotypes with chromosome structure, the use of mutants of Neurospora, bacterial and viral genetics, all described in relatively simple terms.

The chemistry of desoxyribonucleic acid (DNA) and its suitability as a transmitter and storage substance for genetic material is lucidly presented together with the Watson-Crick hypothesis and salient experimental evidence in its support.

A rather extensive presentation of protein structure and relation to biochemical activity is found, and while somewhat tangential to the general theme, should nevertheless be helpful to the layman interested in the chemical basis of the transfer of genetic information from DNA to the proteins which execute the orders. The way in which the four repeating units of the DNA polymer might ultimately code the sequence of a protein polymer containing 20 repeating units is unknown, and several suggested codes are described. Valuable references are given with each chapter.

The step from genetic coding to the coding of the immunological response is a big one and any connection with the organic basis of memory is even greater. Yet one already feels the strong influence of genetic advances on concepts of memory mechanisms. Both desoxyribonucleic acid and ribonucleic acid are frequently mentioned today as possible storage places of the memory trace in the nervous system.

While our knowledge in this latter area is for the moment meager, and it would be foolhardy to make rigorous analogies of the various biological information storage systems, the impressive gains in our knowledge of genetic coding and information storage serve a valuable function in behavioral areas by strengthening the conviction that it will ultimately be possible to find organic correlates and by supporting hope that the brain itself will yield its secrets to the tools of modern biology.

BERNARD W. AGRANOFF
Mental Health Research Institute
The University of Michigan

L. A. Hiller and L. M. Isaacson. *Experimental music.* New York: McGraw-Hill, 1960.

IN this book, the authors, a musician and a mathematician, have programmed various musical rules on a digital computer, and have transcribed a random sample of the printout for string quartet. Christened the *Illiac Suite* in four movements and a coda, it has been played by the Walden Quartet at the University of Illinois. The score is given in an appendix.

Four sets of experiments were conducted. In the first three, rules in the form of test routines are performed on the output of a random number source. In E_1 beginning with a simple cantus firmus three to twelve notes long, two, three, and four voices of

TABLE 1

TABLE OF FUNCTIONS FOR THE GENERATION OF MARKOFF-CHAIN MUSIC IN EXPERIMENT FOUR

Interval	Stochastic Variable v_j	Harmonic Function x_j	Proximity Function y_j	Combined Function $z_j = x_j\, y_j$
Unison	0	13	13	26
Octave	12	12	1	13
Fifth	7	11	6	17
Fourth	5	10	8	18
Major Third	4	9	9	18
Minor Sixth	8	8	5	13
Minor Third	3	7	10	17
Major Sixth	9	6	4	10
Major Second	2	5	11	16
Minor Seventh	10	4	3	7
Minor Second	1	3	12	15
Major Seventh	11	2	2	4
Tritone	6	1	7	8

$\sum_{j=0}^{12} x_j = 91$ $(x_j = x(v_j))$ $\quad \sum_{j=0}^{12} y_j = 91$ $(y_j = y(v_j))$ $\quad \sum_{j=0}^{12} z_j = 182 = 2 \times 91$

simple monody are added. In E_2 each of the rules of Sixteenth Century counterpoint were introduced, one by one, into the computer to produce four-part counterpoint of the first species. Rhythm and dynamics are added in E_3.

In experiments 1–3, each note either meets or fails the tests; failing, it is discarded and the next random number is tested. Not so in the Markov-chain music of E_4, where constraints are imposed on the source. These constraints are an ordered list of intervals. Each interval has a weighting on two functions: a harmonic function and a proximity (melodic) function, relative to some origin pitch. The authors find this distinction in esthetics from Pythagoras to Hindemith. The weights are arbitrarily made arithmetic, as shown in Table 1. The list is ordered in the computer from largest to smallest weight. Each note is formed by taking a random integer R, where $(0 \leqq R \leqq w - 1)$ and adding to $-R$ the weights w_j in sequence until $-R + \sum_{j=0}^{R} w_j$ became positive, whereupon R became that interval associated with the last added weight. Thus no notes need be rejected as in experiments 1–3. And Markov-chain music is far easier to program.

Obviously, the playful possibilities of this scheme are limitless. Every note might be governed by the composite function, or certain beats, parts, or measures may be governed by each of the separate functions By altering the weights in Table 1, the programmer may militate for or against, say, repeated intervals; he may produce pure tritone music or pure unison music, or pure random music if all weights are made equal. The coda of *Illiac Suite* was actually produced by reducing all weights by 1 every three measures, constraining the music gradually to a unison close. This example is most interesting, since some anticipated goal point was allowed to govern the weights determining the path by which it would be reached.

In a final chapter the authors commend the computer as a multipurpose tool in music. In teaching, it may generate exercises using the rules of counterpoint singly or in any combination, something hard for a human. In musicology, it may help to discriminate styles and authorship, similar to the work of Yule and Herdan with languages. In composition, it may extend and change present musical structures. However, many of these hopes will depend on how much we can reasonably expect the statistics of interval distributions to tell us about the syntax of music. Macrostatistics of a natural language are compatible with almost any grammar, or for that matter, with nonsentences made from the word stock, and therefore they shed little light on structure. And attempts to construct statistical approximations to text will simulate text only in the trivial case where enough text is put in bodily to resemble itself. Most of Hiller and Isaacson's "syntactic" constraints are of this sort. If one wants to model a piece of, say, jazz, one can govern the downbeat every eight beats by a harmonic function so weighted that it will give the desired harmonic skeleton of jazz. Furthermore, "self-embedding" constructions which pose insurmountable difficulties in *principle* to statistical simulation, seem to me to be even commoner in music than in natural language. There is likely a whole group of problems which, while present in language, are most prominent in music. Study of these might help avoid some blind alleys already covered.

Though I have not heard the *Illiac Suite* performed, I have studied it at the piano, but—disturbing thought—is it necessary? The suite is selected randomly, not esthetically, from the mass of total printouts. Surely it will be necessary to listen as we program; to make some intuitive judgment to distinguish better from poorer efforts, just as we can do with sentences. After all, a computer's point is its ability to derive a great number of experimental consequences from the same underlying model. This is something we rarely encounter naturally in the arts, and it is one big obstacle to comparisons in these fields. Only by studying the structural secrets of better productions by contrast with their poorer siblings can we hope to guide the choice of what is worth programming.

WILLIAM PAUL LIVANT
Mental Health Research Institute
The University of Michigan

COMPUTERS IN BEHAVIORAL SCIENCE

Publication of this department is partially supported by a grant from the National Science Foundation to the chairman of its editorial committee, Steven G. Vandenberg.

Preparing Psychophysiologic Analog Information for the Digital Computer,[1]

Herbert Zimmer, University of Georgia.

PROBLEM

The excessive demands on time and labor made by the manual and semiautomatic methods of preparing psychophysiologic information for statistical analysis made it seem advisable to explore means of processing such information by digital computer.

An early problem demanding a solution was the determination of the optimal means of translating analog channels into digital information, as well as the best location in the system for accomplishing this. The possibilities were narrowed down to two major approaches: (a) the immediate, and necessarily separate, digitization of each individual channel, or (b) the successive sampling of all channels by an electronic multiplexer, and the transformation of their respective voltages into digital information by a single analog to digital (A–D) converter. In view of the considerable increase in cost of using separate pulse counters and digital voltmeters, a decision was made in favor of the second approach.

It quickly became apparent that the conventional electronic methods which are used to prepare psychophysiologic information for oscillograph presentation would not always be adequate for digital recording. The information must first be edited to suit the peculiar characteristics of the data processing system. Specifically, the voltage of a given channel available to the converter whenever the multiplexer happens to sample it must be representative of the activity of that channel for the time period subsumed under the sample.

Most experiments are addressed to more or less specific questions. Usually, some selected aspects of the psychophysiologic data will suffice to provide the information pertinent to these questions. The rest, and frequently this may be the bulk, of the data need not be scrutinized. Therefore, the aim is to devise the best means for extracting from the total body of the data those factors which are of interest at the moment. Here again, there are basically two alternatives: (a) the condensation and integration of data by analog equipment prior to its conversion into digital format, and (b) a full digital description of the psychophysiologic variable, with subsequent data reduction through appropriate digital computer programming. Whenever appreciable development work is required to put the first alternative into operation, it seems wise to employ the second approach, until the desired refinements can gradually be incorporated.

PSYCHOPHYSIOLOGIC SIGNALS

1. Skin resistance

Two zinc-zinc sulfate electrodes, 2 cm. in diameter, and mounted in lucite cups, are attached to the volar surface of the subject's left metacarpus and forearm. The subject (with 50 μamps. through him) is placed in the feedback circuit of a constant current amplifier. A differential amplifier converts the voltage drop across the subject into a voltage proportional to the skin resistance.

In order to minimize nondeterminable errors due to the polarization of the electrodes, the direction of current through the subject is reversed at 0.5 second intervals by a switching diode, driven by a multivibrator.

[1] This research was supported in whole or in part by the United States Air Force under Contract No. AF 49(638)-487 monitored by the AF Office of Scientific Research of the Air Research and Development Command.

Substantial contributions to the development of this system were made by E. S. Sherrard, H. L. Mason, and C. L. Bolen.

An absolute value circuit serves to rectify the artifactual phasing thus produced, and to maintain a unipolar voltage at the output of the differential amplifier.

Stepping ranges which subdivide the subject's total response range are not required here, since the system can record a response as small as one part in 1,000.

2. Respiration

The transducer for respiration is a single-turn, linear, precision potentiometer forming one arm of a D.C. bridge. One transducer is suspended about the subject's sternum and another slightly below his diaphragm. Bands encircling the subject and spring returns move the potentiometer arms. Differential D.C. amplifiers boost the voltages across the bridge to recording level. Respiration is sampled every 0.1 second, to yield a detailed description of the respiration cycle, without analog editing for this channel.

It is possible to edit both respiration amplitude and respiration cycle duration prior to conversion, through the use of peak followers, zero-crossing detectors, integrators, and resetting relays. Such an approach would lead to considerable complexity for this channel, would yield less information, and could require a sampling rate as fast or faster than the unedited respiration signal with no consequent saving in digital computer time.

3. Muscle action potential voltage

Two fine silver discs, 5 mm. square, are attached with adhesive tape over the frontalis muscles, with a distance of 5 cm. between them. The approximately one million amplification required for this signal is provided by two cascaded, low-level, differential A.C. preamplifiers.

To obtain readings of the total muscle action potential voltage, the signal is first rectified and then integrated, in order to avoid a very high sampling rate. The integrator is sampled and reset every 0.1 second by the end of the conversion pulse from the A-D converter.

4. Heartrate

A modification of ECG lead II is employed. A 3.1 by 5.2 cm. stainless steel electrode is strapped over the left tibialis anterior muscle, and a fine silver disc of 19 mm. diameter is fastened to the right ear lobe. A Fels Cardiotachometer (Yellow Springs Instrument Company) converts the interval between *R* spikes to an analog voltage, which is proportional within 1 % to this time interval.

5. Finger pulse amplitude

The transducer is a 12 by 1.5 by 0.5 mm. piece of piezo-electric material taped to the volar surface of the terminal phalangeal joint of the third finger of the left hand. This signal is picked up by a differential A.C. amplifier with a variable gain of 100 to 1,000. A peak follower provides a prolonged reading of maximum pulse amplitude. The reset signal is obtained from the negative phase of the first ECG *R* spike which follows the pulse wave. Maximum pulse amplitude is thus available to the A-D converter during most of the interpulse period.

6. Skin temperature

A thermistor probe, forming one leg of a D.C. bridge, is taped firmly over the left risorius muscle. The output voltage of the bridge has an inverse, though linear, relation to skin temperature.

7. Body-weight shifts

Pneumatic cushions under the seat of the subject respond to shifts in body weight and produce resistance changes in a pressure transducer. This signal is rectified, integrated, sampled and reset at 0.1 second intervals.

The use of two channels of both respiration and muscle action potential brings the total number of channels to nine. The ten-channel sampling accommodates up to ten analog channels, but works as well with fewer active inputs.

RECORDING

Sampling switch

A ten-channel sampling switch facilitates the sampling of each of ten channels at 0.1 second intervals: a total of 100 samples per second. A bank of ten operational amplifiers adjusts the full-scale signal level of all

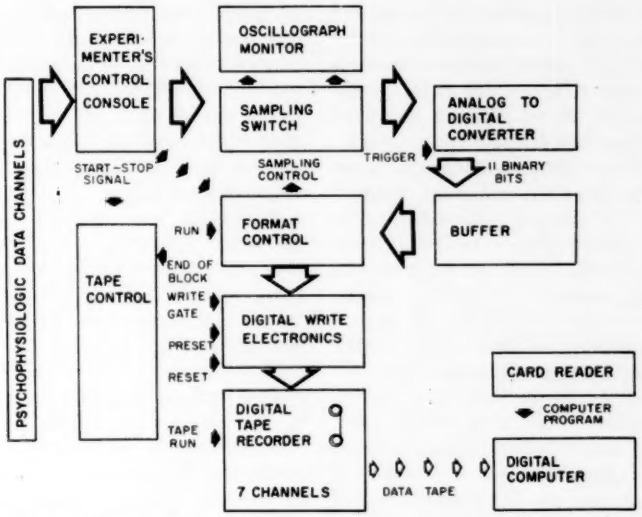


FIG. 1. Data Flow Chart. The open arrows indicate the progress of psychophysiologic information through the system. The solid arrows show internal control signals of the system.

channels to approach ± 10 volts, the full-scale range of the A-D converter. Ten analog and-gates, controlling electronic switches, are activated in sequence by pulses from character and word counters in the format control unit of the system and cause the operational amplifier outputs to sequentially feed the A-D converter input. The output of these and-gates is delayed, amplified, and inverted to provide the trigger signal for the A-D conversion of each of the psychophysiologic channels.

Buffer amplifier

In order to make the eleven binary bit output of the A-D converter acceptable to the format control unit of the system, the buffer produces +20 volts at the output for a binary one, and 0 volts for a binary zero.

Format control unit

Digital data from the converter and buffer are organized by this unit into format for non-return-to-zero (NRZ) recording on seven-channel magnetic tape. The diode selector matrix permits the recording of digital data from other sources in addition to the A-D converter, and has a capacity of eight 36-bit words, that is, 48 six-bit character cells. It is controlled by an electronic scanning system consisting of six character drivers, eight word drivers, seven line drivers, a character ring counter, a word ring counter, a character multivibrator, a word multivibrator, a beam former, a time base generator, and clock control. The diode matrix acts as a commutator in sequentially connecting six output lines to groups of six inputs.

Digital write electronics

A preset signal from the tape control unit is employed to have the head current flow in the reference direction, when the write gate is energized. Once the write gate is activated, the flip-flops accept inputs from the format control. The flip-flops reverse head current whenever a binary one appears at their input. Seven transistorized flip-flops, which supply sufficient magnetizing current to the write heads to saturate the tape, make up the write electronics.

Tape

The tape runs continuously at a speed of 1″ per second, and has a density of 200 characters per inch, resulting from 100 eleven-bit conversions per second. Psychophysiologic reactions associated with a single stimulus are enclosed by ¾″ gaps, encom-

passing a 30″ time interval. A ten-second prestimulus baseline is interrupted by the occurrence of the stimulus, some milliseconds in duration, and followed by a 20″ poststimulus reaction period. At the indicated speed, a 2,400 foot reel of tape can accommodate eight hours of continuous recording, or about 15 half-hour experimental sessions together with identifying information. The information stored on the tape is in binary code and IBM format, and can be processed on any digital computer which can accept this format.

With three exceptions (the Brush oscillograph monitor, the Epsco Datrac 611 A-D converter, and the Potter 3280 tape transport), all components of the A-D conversion system were built by the project technicians and engineers after the specifications and with the guidance of the National Bureau of Standards. The replacement cost of the system can only be estimated: $50,000 to $80,000.

It is hoped that a system of this sort will foster more extensive psychophysiologic research by making the process of data analysis less cumbersome, less time-consuming, and more accurate.

(Manuscript received October 27, 1960)

Automated Programming for Multinomial Distribution Problems on the IBM 650, *James E. Farmer, Washington State University.*

Many problems which require analysis of multinomial distributions originate from questionnaire or survey type data. The social sciences (business, marketing, economics, psychology, sociology, etc.) make extensive use of questionnaires or surveys to obtain information as a basis for decision making or problem solving. Often these solutions or decisions depend in large part on the distribution of the basic information.

Methods for obtaining multinomial distributions vary from hand tally methods through card counting sorters and tabulators to electronic computers. The use of electronic computers for obtaining distributions provides three very desirable advantages over other methods: (1) accuracy, (2) greatly shortened analysis time, and (3) the results are in suitable form (punched cards or tape) for further analysis by automatic methods. In addition, when furnishing the count for each cell of a distribution the computer can at the same time calculate the percentage-of-the-total associated with the count of each cell.

Except in the area of large distribution problems the computer generally has not been the most economical method for obtaining distributions. The major obstacle to economic use of a computer is the programming time requirement. However, an inefficient program can also be a factor. A system which minimizes personnel time and maximizes computer efficiency makes the computer a very economical method for obtaining distributions in all but small problems.

Various methods for obtaining solutions to distribution problems utilizing an IBM 650 computer with index registers are possible. Certain methods for the solution of the problem are enforced by the characteristics of a magnetic drum memory. For the most part, surveys or questionnaires involve responses to a large number of questions, with a varying number of possible responses to each question. The analysis of all questions in one operation is an ideal computer application and a relatively straight-forward problem.

Alternate methods which may be considered for the solutions to distribution problems are: (a) programming each problem separately by hand methods, (b) one general distribution program to be used on all problems, and (c) a program generator which provides a different program for each problem.

Individual programming of separate distribution problems may or may not provide efficient operating programs, depending on the programming techniques employed. A distribution program has a large number of repetitive steps. On a drum memory computer, loops for the repetitive portion of the program seriously handicap the program operating speed. If loops are not used a great deal of programming time is needed. Programming separate distribution problems

thus offers the potential advantage of efficient computer use but has the disadvantage of requiring more programming.

A general distribution program can be defined which will provide solutions to most distribution problems. For small memory computers the definition of such a general distribution problem presents a dilemma. How should the proportion of one and two digit responses be divided within the available memory to provide the best use of the memory for the largest number of problems? From the standpoint of machine time use, division of a word to store more than one count is not a feasible method for increasing available memory. A general distribution program would operate efficiently only for problems closely resembling the model problem, which would fit only a portion of the problems. The major disadvantage of a general program is efficiency in running. However, elimination of programming for each problem is an advantage.

A generator program produces as output an entirely different program, generally referred to as the object program. The object program obtains the solution to one particular problem and a new program is generated for each different problem. The object program will vary as a function of the parameters used to define each problem. A distribution generator possesses the advantages of both previous methods. Time requirements for operation of generator plus object program should be comparable to the time required for operation of a general distribution program. Computer efficiency of an object program produced by the generator will be better than the efficiency of a general program, and equally efficient when compared with individual programs. The machine time requirement for program generation should be less than, or at worst equal to, the time requirements for checking out individual programs.

The generator method was chosen by the author as the best method for obtaining solutions to distribution problems. A Distribution Program Generator has been coded, and is currently being utilized by the Computing Center at Washington State University, which "writes" a program to solve a specific distribution problem, within the following limitations: (1) number of observations $< 10{,}000$; (2) number of items or questions < 81; (3) largest numerical response to any item < 100; (4) total number of responses < 1001. Larger problems may be partitioned and handled as more than one problem.

Input to the generator is in the form of control cards which specify the total number of items for which distributions are to be produced and the highest numerical response for each item (three control cards maximum). From this information the generator computes and assigns storage areas for counting purposes, assigns the base addresses required for obtaining these counts, and punches out the object program.

Output from the generator is a one instruction-per-card symbolic program in SOAP II format. The symbolic program must be assembled using SOAP II. An assembly phase is essential to efficient program operation, since the optimum location for instructions is dependent on the parameters of each distribution program. The assembled distribution program is used to prepare the desired distributions.

The original data cards punched from the questionnaire or survey are used as input to the distribution program, because specific arrangement of the items for which distributions are needed is accomplished through board wiring, which is dependent upon each problem. Two-way distributions are easily prepared by sorting the input data on one of the desired classifications. For example, if the distribution for all questions were desired by income groups, the input would be sorted by income groups. With appropriate board wiring, a separate distribution would be punched for each income group.

Output from the distribution program consists of one card for each 10 responses for items having 11 or more responses and one card per item for items having less than 11 responses. The information given for each response is the frequency count for each response, and the percentage of this count to the total number of observations. In addition, each card contains identification and

the total number of observations. The fixed nature of the output allows a permanently wired tabulator panel for listing.

An idea of the over-all effectiveness of the generator program may be obtained from the time required for various phases of the operation. Generation and assembly will take from 4–6 minutes for small problems to 18–20 minutes for maximum-size problems. The distribution program requires approximately 0.3 seconds computing time per observation for 12 items, and 1 second per observation with 80 items to be distributed. Output time for the largest distribution program is about 1.5 minutes per distribution. A problem solving 1200 observations on 40 items was run in 40 minutes.

The Distribution Program Generator makes the computer economical for obtaining distributions in all but small problems. Accuracy and punched card results are additional assets.

The program and its description are available from the Washington State University Computer Center.

(Manuscript received November 29, 1960.)

COMPUTER PROGRAM ABSTRACTS

IBM 704 Program for General Analysis of Variance, *Allan W. Dickinson, Monsanto Chemical Company, St. Louis, Mo.* **(CPA 39)**

Description: The program performs analysis of variance computations for up to twelve factors at any number of levels. Sums of squares, degrees of freedom, and mean squares are computed for all main effects, first and second order interactions, and the residual. Components of variance for all main effects or for selected main effects are available on an optional basis. Also printed are tables of means for main effects, first and second order interactions. Any degree of multiple or fractional replication can be handled. Standard transformations on the input data (log, square root, etc.) are available.

Computer: IBM 704, 8K core, 1 tape (optional). **Program Language:** Fortran II. **Comments:** Sums of squares are computed in single precision floating point arithmetic. If greater accuracy is desired, the computer can be instructed to perform the analysis on deviations from means rather than from the raw data.

A choice between printer output and tape output is available through a sense switch control.

If there is sufficient interest a manual for this program will be made available by the American Institute of Chemical Engineers, 25 West 45th Street, New York 36, N. Y. Publication of this abstract is by arrangement with the A.I.Ch.E. Reprinted from the April, 1960 issue of *Chem. Eng. Prog.* where it was program 047.

Datatron 205 Program for Polynomial Equation Fitting, *W. L. Dieckmann, Mathematical Services Group, California Research Corporation, Richmond, California.* **(CPA 40)**

Description: This program computes the coefficients of a polynominal equation of the form: $Y = C_0 + C_1X + C_2X^2 + C_3X^3 + \cdots C_nX^n$ given a number of data points equal to $n + 1$. Equations as high as the nineteenth degree can be handled. The program is not restricted to equal intervals for the data points. If desired, back-solutions for other values of X may be obtained.

Computer: Datatron 205, 4000 words storage, card input and output, and floating point unit. **Program Language:** Datatron 205 Machine Code. **Running Time:** Running time depends upon the degree of the polynominal equation. For a fifth degree equation (including back-solutions), about four minutes is required.

A manual will be prepared for publication by the American Institute of Chemical Engineers, 25 West 45th Street, New York 36, N. Y., if sufficient interest develops. Publication of this abstract is by arrangement with the A.I.Ch.E. Reprinted from the August, 1960 issue of *Chem. Eng. Prog.* where it was program 033.

IBM 704 Program for the Solution of Simultaneous Linear Equations, *D. J. Kaufman and C. W. Woo, Monsanto Chemical Company, St. Louis, Missouri.* **(CPA 41)**

Description: The program will read in non-zero elements of the $n \times (n + 1)$ matrix of coefficients and right hand sides for a set of simultaneous linear equations. These are solved by means of a Gauss reduction with error iteration and the answers are then printed out. It is possible to save the original matrix and merely read in changes to the original set of equations. Input and/or output may be on magnetic tape, if desired. On an 8K 704, up to fifty equations may be handled. **Computer:** IBM 704 (8K). **Program Language:** 704 Fortran II. **Running Time:** Less than one minute for sets of up to fifty equations.

A program manual may become available from

the American Institute of Chemical Engineers, 25 West 45th Street, New York 36, N. Y., if sufficient interest develops. Publication of this abstract is by arrangement with the A.I.Ch.E. Reprinted from the September, 1960 issue of *Chem. Eng. Prog.* where it was program 060.

Burroughs 205 Program for Selective Curve Set, *J. R. Walker, Celanese Chemical Company, Data Processing System, P. O. Box 561, Corpus Christi, Texas.* (**CPA 42**)

Description: Given a set of x-y data, this program calculates the coefficients for thirty-six simple, arbitrarily selected empirical equations. Three additional quantities are calculated for each equation to help determine the quality of the fit.

The thirty-six equations are obtained by analyzing the dependent variable y for a relationship to the independent variable x as:

$$y = m\phi(x) + B$$

$$y = \frac{1}{m\phi(x) + B}$$

$$y = 10^{m\phi(x)+B}$$

where $\phi(x)$ is sequentially: x, $1/x$, $\log x$, e^x, x^2, $1/x^2$, $x \log x$, $(1 + x)x^2$, $(1 - x)x^2$, $\log x/x$, $(1 + x)^2$, $(1 - x)^2$. Each equation's coefficients are calculated by the method of zero moments. A standard deviation, which is the standard deviation of the residues from the predicted line, the maximum absolute deviation that occurs in the data set, and the position of the maximum deviation are reported for each equation.

Computer: Burroughs 205; paper tape input-output. **Program language:** Burroughs 205 machine code. **Running time:** The running time is dependent on the number of data points used. A sample run containing 16 points ($32x - 6$ values) required approximately one minute per equation including punch-out of the calculated coefficients and constants along with the alphabetic equivalent of each equation.

A manual may be ordered from the American Institute of Chemical Engineers, 25 West 45th Street, New York 36, N. Y. Publication of this abstract is by arrangement with the A.I.Ch.E. Reprinted from the July, 1960 issue of *Chem. Eng. Prog.* where it was program 053.

IBM 650 and Burroughs 205 Programs for the Design of Optimum Multifactorial Experiments, *W. N. Smith and A. W. Umland, The Dow Chemical Company, Process Control Laboratory, Freeport, Texas.* (**CPA 43**)

Description: The program is used for the statistical design of experiments and/or the analysis of data from these experiments. Either function can be performed separately. For a given number of independent variables, n, to be evaluated by making N runs, a two-level orthogonal design is determined. N must be greater than n and some multiple of four ranging from 8 to 96. The results of such a design can be analyzed, listing the effect of changing each independent variable. Those variables found statistically significant are listed separately. The changes in the independent variables that give the maximum rate of increase of the dependent variable (path of steepest ascent) are readily obtained from the analysis. The program is based on the method of Plackett and Burman (1946).

Computer: This program (with some modifications) has been written for a Burroughs 205 and also for an IBM 650. The 205 has floating point unit, card reader and on-line printer. The 650 is a basic unit. Fix-float and square root subroutines are used. An f-table subroutine is used with the 205 program only. (The 650 program ends with the calculation of the variable effects, variable mean squares, and error mean squares.)

Program language: The 650 program is in SOAP I. The 205 program is in machine language. **Running time:** Running time varies with the size of the design. On the 205, the program requires 99 seconds to design a 16-run experiment and 61 seconds to analyze the results from 16 runs. The time on the 650 is somewhat longer. **Comments:** In the analysis of variance, either 5% or 10% f-table subroutines are available. Also, any number of the variables can be pooled with the error mean square.

Publication of this abstract is by arrangement with the American Institute of Chemical Engineers. It is reprinted from the November, 1960 issue of *Chem. Eng. Prog.* where it was program 064.

Reference:

Plackett, R. L. and Burman, J. P. Design of optimum multifactorial experiments. *Biometrika*, 1946, 33, 305–325.

Least Squares Fit to Relaxation Equation, *P. J. Cislak and A. H. Emery, Purdue University, Lafayette, Indiana.* (**CPA 44**)

Description: This program calculates the values b' and c' which give a least squares fit to the equation, $y = b'(1 - ae^{-x/c'})$. The method used is a Newton-Raphson fit to determine c', up to a given tolerance, with the restriction that the squared deviations be minimized. The program will handle up to 800 observations.

Computer: Datatron 205, 4000 words storage,

card input and output, floating point unit. **Running time:** Approximately five minutes for twenty observations. This time is dependent on the number of observations and the initial approximation of c'.

Publication of this abstract is by arrangement with the American Institute of Chemical Engineers. Reprinted from the November, 1960 issue of *Chem. Eng. Prog.* where it was program 050.

Reference:

Hildebrand. F. B. *Introduction to numerical analysis.* New York: McGraw-Hill, 1956, p. 450.

Minimal IBM 650 Program for Numerical Solution for Certain Systems of Differential Equations, *N. E. Welch, D. S. Billingsley, and C. D. Holland, Texas A. & M., Dept. of Chemical Engineering, College Station, Texas.* **(CPA 45)**

Description: Using the Runge-Kutta method, this program solves the following system of equations:

$$(1) \qquad \frac{AdW}{dZ} + \frac{CdX}{dX} + W = G$$

$$(2) \qquad \frac{dX}{dZ} = \left[\exp\left(B - \frac{B}{W}\right)\right](1 - X)f(X)$$

where $f(X)$ has one of the following three forms:

$$f(x) = 1$$

$$f(X) = (1 - X)$$

$$f(X) = [1 - Y(1 - X)]$$

and where

$$0 \leqq X < (1 - Y)/Y;$$

$$0 < W \leqq G; X < 1$$

Restrictions on the parameters A, B, C, G, and Y are discussed below. This system of equations with $F(X) = 1$ has been used by Billingsley, et al. (1958), to describe certain systems in reactor design involving the first order reaction $A \rightarrow$ Products. Similar systems involving the second order reaction $2A \rightarrow$ Products or $A + B \rightarrow$ Products may be described by this set of equations when $f(X) = [(1 - X)]$ or $f(X) = [1 - Y(1 - X)]$ respectively.

Since all computations are in fixed point, arithmetic overflows in the accumulator should be expected when data do not lie within the following limits:

$$0 \leqq C \leqq 0.18$$

$$0.005 \leqq A \leqq 0.111$$

$$5 \leqq B \leqq 55$$

$$1.02 \leqq G \leqq 1.18$$

$$0 < Y < 1.0$$

Among other ways, the program was checked using $f(X) = 1$ and combinations of the above limits which appeared most likely to result in overflows. These limits were used since they were the ones of interest in reactor design. It is entirely possible that the program will handle certain problems where the data exceed some of these limits.

Computer: IBM 650, no accessories other than the read-punch unit. 1000 word drum. The program occupies locations 0000–0354. **Program Language:** Soap I. **Running time:** Approximately 7 minutes for a problem where $f(X) = 1$, $\Delta X = 0.01$ and the initial value of X is zero.

A limited supply of copies of the program are available and may be obtained from the Data Processing Center, Texas Engineering Experiment Station, College Station, Texas, by requesting Program No. A.M. 940. A complete program manual will be made available by the American Institute of Chemical Engineers, 25 West 45th Street, New York 36, N. Y., should sufficient interest develop. Publication of this abstract is by arrangement with the A.I.Ch.E. Reprinted from the May, 1960 issue of *Chem. Eng. Prog.* where it was program 051.

Reference:

Billingsley, D. S., McLaughlin, W. S., Welch, N. E., Holland, C. D. Use of computer in reactor design. *Ind. Eng. Chem.*, 1958, 50, 741–752.

An IBM 650 Program for Hierarchical Analysis of Variance. HI-AOV, *Gene Pulley, Oklahoma State University* **(CPA 46)**

This program is designed to compute the analysis of variance for a hierarchical classification design having five stages or less. It also computes the coefficients of the variance components for each mean square. The program will handle as many as ten variables simultaneously, each variable being limited to five digits or less. The mathematical model for each variable may be written as:

$$Y_{ijklm} = \mu + A_i + B_{ij} + C_{ijk} + D_{ijkl} + E_{ijklm}$$

The IBM 650–653 with the following hardware is necessary: Sixty words immediate IAS core

storage; Three index registers (A, B, and C); Floating point devices; IBM 533 read-punch unit and proper control panel.

This program does not require a header card, since all controls are held within the program. All variable arithmetics are done in double-precision, fixed-point arithmetic. A maximum of 100,000 entries on each of the ten variables is provided for, without possibility of overflow. Missing data that is properly identified will not disrupt the processing. This, along with complete validity and sequence checking, makes HI-AOV (Hierarchical) virtually fool-proof. Since machine time may be restricted to short intervals, a provision is made which will allow the operator to leave the machine any time during processing and complete the processing at a future time. Request for this program should be made through a university or other recognized computer center and should be addressed to Computing Center, Oklahoma State University, Stillwater, Oklahoma.

"Seeing, then, that even sense impressions not only can but must be rationalized in irreconcilably different ways, according to the class of object with which they are supposed to be connected, we need not be troubled or surprized by the contradictions which we find in the rationalization of less uniform psychological experiences. Thus, the almost indefinitely numerous rationalizations of the aesthetic and the mystical experiences not only contradict one another, but agree in contradicting those rationalizations of sense experience known as scientific theories. This fact greatly disturbed our grandfathers, who kept on losing their faith, sacrificing their reason, striking attitudes of stoical despair, and, in general, performing the most extraordinary spiritual antics, because of it. Science is 'true,' they argued; therefore art and religion, therefore beauty and honor, love and ideals, must be 'false.' 'Reality' has been 'proved' by science to be an affair of space, time, mass, number, and cause; therefore all that makes life worth living is an 'illusion.' Or else they started from the other end. Art, religion, beauty, love, make life worth living; therefore science, which disregards the existence of these things, must be false. It is unnecessary for us to take so tragic a view. Science, we have come to realize, takes no cognizance of the things that make life worth living, for the simple reason that beauty, love, and so on, are not measurable quantities, and science deals only with what can be measured. One psychological fact is as good as another. We perceive beauty as immediately as we perceive hardness; to say that one sensation is illusory and that the other corresponds with reality is a gratuitous piece of presumption."

ALDOUS HUXLEY, *Do What You Will*

ABSTRACTS OF CURRENT LITERATURE WITH INTERDISCIPLINARY IMPLICATIONS

Prepared by Members of the Staff of the Mental Health Research Institute

KUBIE, L. S. **Is preventive psychiatry possible?** *Daedalus*, 1959, 88, 646–667.

This article outlines in the briefest possible way the immense task that lies before us in the application of psychiatry to the prevention of mental illness. Any form of preventive medicine invariably meets obstructive forces that are rooted in socially entrenched prejudices. These forces are even more tenacious in the field of psychiatry because of the tendency of the neurotic processes, which have been incorporated wholesale into our social institutions, to resist change.

Resistance to change, in an individual, is a defense mechanism against uncovering the deeper conflicts that lie beneath the surface. Many human cultural institutions tend to exploit, reinforce, and reward the neurotic component in human nature. In order that we may root out these secondary and tertiary processes, a searching examination would be required of all of our institutions, the family, the political and economic systems, the arts, and religion. While it is problematical that any of these institutions generate neuroses, certainly many of them intensify such trends by actively rewarding them, while others exploit and prey upon them. An unsparing effort to broader application of psychiatric knowledge would challenge every human institution and marshall all the forces of resistance that they are capable of exerting.

However, the basic problem still lies with the primary process which goes back to earliest infancy. The ultimate aim of preventive psychiatry is to limit or reverse the first steps in the neurotic process as these arise in infancy and childhood. At the present time we can recognize several points at which some effort must be exerted. These are: (a) control of those experiences which impose on the infant persisting central emotional positions; (b) control of those conditioning experiences which link these central emotional positions to trigger mechanisms; (c) control of the process of identification by which the developing personality makes destructive and conflicting identifications with destructive and conflicting persons; (d) finally, and most important, there must be a significant measure of control of the process of repression which produces a dichotomy in the symbolic process itself, i.e., the fateful dichotomy into conscious and unconscious components.

None of these goals is easy to achieve; but the recognition of their importance is a step forward. In the coming years the course of preventive psychiatry will depend in part on the development of techniques aimed at these objectives. (*WJH*)

SJOBERG, GIDEON. **Contradictory functional requirements of social systems.** *J. Conflict Resolution*, 1960, 4, 198–208.

Sjoberg amends the existing structural-functional theory of social systems to accord recognition to "contradictory functional requirements." The extension is aimed specifically at interpretations of social change and of recurrent tensions within and between nation states. The mainstream of present theory tends to exaggerate harmony and integration. In contradistinction to this emphasis, another group of theorists have used the conflict or contest model. This article represents a synthesis.

Functional imperatives within all social systems are seen as being plagued by contradiction, thus generating tension and change within the structure. The contradictions stem from both internal and external requirements. Antagonistic structures within a society may promote tension within the system but this should not mean that the

system itself is not being maintained. Sometimes the dysfunction among parts is actually necessary to balance or provide checks upon excesses which might follow from adherence to a few mutually compatible requirements. The following examples are offered.

Industrialization requires hierarchical divisions of labor while simultaneously generating equalitarian values. Deviation from norms is a functional requisite for sustaining a system under conditions when creative solutions are needed. Extremism in politics or prolonged continuance of a garrison state serves to arouse antagonism which may cause a system to oscillate between different sets of requirements. Democracy as a mode of government provides negative feedback thus regulating antagonisms usually before the point of violence. Some such mechanism is required at the international level and the U.N. is seen as a step in this direction. The illustrations support the view that conflict, as well as harmony, inheres within and between social systems. (*MP*)

BITTERMAN, M. E. **Toward a comparative psychology of learning.** *Amer. Psychologist,* 1960, 15, 704–712.

Beach's finding of ten years ago that approximately 90 percent of all animal studies are done with the rat still holds today. This trend can be thought of as the result of a Darwinian approach which assumes that all psychological (as well as structural) phenomena are on a continuum so that rat learning may be different in degree but not in kind from human or inframammalian learning.

Superficial support of the Darwinian hypothesis has been supplied by Skinner with cumulative frequency curves for pigeons, rats, and monkeys trained on a multiple fixed - interval - fixed - ratio reinforcement schedule, and by various maze investigators using the Hampton Court maze with sparrows, rats, and monkeys 55 years ago, both of these showing apparent similarity across species.

The Darwinian hypothesis should not be accepted on superficial data of this sort. There have been attempts by Yerkes and by Harlow to do true comparative studies. These have been very long series of experiments dealing with performance or changes in performance in standardized situations.

There is, however, a different kind of comparative research. Taking the rat as a point of departure, a different species of marked phylogenetic contrast is studied under analogous conditions. The two animals are not compared in terms of absolute scores. Rather, the goal of this approach to comparative psychology is to formulate a "theory of the other animal" with which to compare the "theory of the rat." While such an approach may make "control by equation" difficult, there is the possibility of "control by systematic variation." An example is the study of fish behavior under various reinforcement regimes in a situation adapted to the capacities of this species. Interestingly enough, it has been found that the fish extinguishes more rapidly under partial reinforcement schedules than under constant reinforcement, thus behaving more closely than the rat in accordance with predictions of most current Stimulus-Response theory.

Work with little-studied species requires both the development of techniques suitable to the species and the discovery of laboratory conditions conducive to the maintenance of the animal's good health. Techniques are described for the study of learning in fish, crab, blowfly, and earthworm. (*GR*)

CAMPBELL, DONALD T. **Blind variation and selective retention in creative thought as in other knowledge processes.** *Psychol. Rev.*, 1960, 67, 380–400.

The author discusses the possibility of viewing creative thought in a manner similar to the schemes applied to the lower knowledge or evolutionary processes. The scheme, of course, consists of the concept of blind - variation - and - selective - retention. Campbell asserts that creative thought implies that there are continual "break-outs" from the limits of available knowledge. The concept of "blind" variation, he asserts, does not imply equiprobability or statistical independence between one variation and the next. Indeed, processes involving a system-

atic scanning sweep are recognized as "blind" because variations are produced without prior knowledge of which ones will be successful. Quoting Ashby, he states that what are needed for the blind-variation-and-selective-retention mechanism to work are: (i) a mechanism for introducing variation, (ii) selection by the somewhat consistent or knowable vagaries of the environment, and (iii) preservation and duplication of the selected variations.

The author then quotes various writers whose conception of the higher mental processes agrees essentially with the blind-variation - and - selective - retention notion (e.g., Brain, Mach, Poincaré, Woodworth, Tolman, Schlosberg, etc.).

In commenting on the criticism which has been leveled against blind-variation-and-selective-retention as an explanatory model for higher thought processes, the author makes use of the arguments employed to defend evolutionary theory itself. Thus he argues that in biological evolution and in thought, the number of variations explored is greatly reduced by having selective criteria employed at every step; that much of organic evolution and creative thinking is multipurposed and variations produced need not be fitted to only one end at any given time. As to the *Individual Differences and Genius* argument which is brought to bear against the blind variation concept, Campbell shows that his concept can deal readily with these phenomena. Campbell concludes by admitting that while he believes the evolutionary concept to be useful for understanding creative processes, the fine details of the mechanisms involved are as yet quite sketchy as compared to the state of knowledge attained in biology. (*JWG*)

FISHMAN, V. AND GOLDENBERG, H. **Metabolism of chlorpromazine: Organic extractable fraction from human urine.** *Proc. Soc. Exptl. Biol. & Med.*, 1960, 104, 99–103.

Despite the widespread clinical usage of chlorpromazine, considerably more is known about its metabolic fate in the dog and the rat than in man. In the present paper, the authors have investigated those chlorpromazine degradation products extractable from alkaline solution by ethylene dichloride. Nine compounds were found in this fraction which appeared to be uniquely related to phenothiazine administration. Of these, six appeared to be sulfoxides as judged from their ultraviolet absorption. Chlorpromazine, chlorpromazine sulfoxide, desmethyl chlorpromazine sulfoxide (lacking one methyl group on the side chain), desdimethyl chlorpromazine sulfoxide (lacking both methyl groups), and desdimethyl chlorpromazine were tentatively identified and the desmethylated sulfoxides appeared to be the major extractable degradation products. These results differ both qualitatively and quantitatively from animal studies.

These extractable materials apparently represent only a portion of chlorpromazine metabolites, others remaining in the aqueous phase some of which appear to be phenolic derivatives linked to a glucuronic acid moiety.

A point of considerable interest is an apparent difference in individual excretion patterns, a difference which may be related to the variable clinical efficacy of phenothiazine. (*AY*)

The simplest rudiment of mystical experience would seem to be that deepened sense of the significance of a maxim or formula which occasionally sweeps over one.

WILLIAM JAMES, *Varieties of Religious Experience*

ABOUT THE AUTHORS

Robert F. Fagot ("A Model for Equisection Scaling") is a previous contributor to *Behavioral Science* with his article on "A Model of Riskless Choice" in the January, 1959 issue. Assistant professor of psychology at the University of Oregon, Dr. Fagot continues his interests in mathematical models in psychology, psychophysics, and measurement theory with his present research on the construction and testing of measurement models in psychophysics. This research, supported by the National Science Foundation, continues along the same lines begun under Professor Patrick Suppes at Stanford. He is a graduate of M.I.T. where he received a B.S. in meteorology and of Stanford University where he obtained his Ph.D. in psychology and statistics.

James E. Farmer ("Automated Programming for Multinomial Distribution Problems on the IBM 650"), an assistant computing analyst at Washington State University, received his undergraduate education at the University of Idaho and an M.S. degree from Utah State University. The author of a number of IBM 650 programs, he is presently developing a program on "Infantry Small Unit Tactical Simulation." Mr. Zimmer writes that his major field of interest is in techniques and methods of generalized data processing for research purposes.

William W. Haythorn ("An Analysis of the Decision-Making Functions of a Simulated Air Defense Direction Center") writes that his present professional interests are devoted to research on the performance of large organizations "particularly with regard to the role of man in managing such systems." Dr. Haythorn, manager of the Logistics Systems Laboratory of the RAND Corporation at Santa Monica, California, received his B.A. degree from West Virginia University and his Ph.D. in social psychology from the University of Rochester. He is the author of several articles and papers dealing primarily with the use of simulation techniques.

Nicolas Malleson ("Instinct and History"), Director of the Research Unit for Student Problems at the University of London Institute of Education, says of his article, "I have been interested in the application of animal studies to human affairs for many years, since first coming across Zuckerman's work on the Social Life of Apes. It has seemed to me that there are interesting and fruitful possibilities if one can devise the right intellectual vehicles for penetrating the 'culture barrier'; this is what I am groping after in this very tentative article." The author of numerous reports on the university student and his problems of adaptation, Dr. Malleson received the major part of his education at Cambridge where he earned a B.A. with honors in 1942 and became an M.A. in 1950. He was graduated with an M.B., B.Chir. (bachelor of medicine and surgery degree), Cambridge, in 1945, receiving an M.D. there in 1957. In 1950 he took the specialization boards in internal medicine and became a member of the Royal College of Physicians. Dr. Malleson spent the year 1952 at The University of Michigan where he was a senior Fulbright Research Scholar, and returned again to the states in 1956 as a delegate to the International Conference on Student Mental Health held in Princeton. Interested in the role that instinct plays in social processes, he believes that "in the next twenty years the development of psychological and psychotherapeutic thought will see an increasingly sophisticated consideration of the role of instinct, and a more favorable reappraisal of the views of many earlier psychologists here, MacDougal, Dreever, and Thorndike for example. In the twenties, the baby was sometimes thrown out with the bathwater."

John E. Overall ("Models for Medical Diagnosis") is chief of Criterion Development, Central Neuropsychiatric Research Laboratory, Veterans Administration Hospital, Perry Point, Maryland. He received his Ph.D. in 1958 from The University of Texas and spent the following year as a National Science Foundation Postdoctoral Fellow at the Psychometric Laboratory at the University of North Carolina. Dr. Overall, whose major research efforts are directed toward the development of quantitative procedures for use in the evaluation of new psychotherapeutic drugs, became interested in the present research through previous work on quantitative classification of psychiatric patients. With regard to this paper, he writes of his collaboration with Dr. Williams since 1957 "when we both were members of an interdisciplinary team at the Radiobiological Laboratory of the University of Texas and the United States Air Force. This collaboration has led to a mutual appreciation of problems shared by medicine and psychology. While I was engaged in the application of multivariate statistical methods to psychiatric problems, Dr. Williams stimulated the present inquiry into the implications of these procedures for the broader problem of medical diagnosis. The present article is an attempt to describe some of the potential applications of available statistical procedures to

problems of medical diagnosis." Dr. Overall is the author of approximately 30 research papers, several in areas related to medicine and biology.

Gerard A. Radnitzky ("Some Remarks on the Whorfian Hypothesis"), a member of the philosophy department at Gothenburg University in Sweden, came to philosophy by way of initial training in psychology, education, and statistics. He writes, "I began increasingly to feel the need for a background in the philosophy of science in order to come to grips with psychological theorizing." After receiving his B.A. degree from Stockholm University in 1954, he met H. E. Törnebohm, at that time acting professor in "teoretisk filosofi" at the University of Stockholm. Dr. Radnitsky, greatly influenced by the teachings and writings of Törnebohm who was originally a physicist and had become a philosopher, went on to study "teoretisk filosofi" (logic, semantics, and philosophy of science) at Gothenburg University where he earned his Filosofi licentiat (Ph.D.) in 1958, and his Fil. lic. in "praktisk filosofi" (mainly ethics) in 1959. He goes on to say, "A considerable practical difficulty with doing metascientific work requires both a good background in the discipline concerned as well as sufficient training in and a suitable method of philosophy of science. Team work appears indicated, but even if it is established there are still difficulties of communication to be overcome until some workable frame of reference has been established. At present, I am working on a book with Carl Lesche (who from physics via biology and psychology came to the philosophy of science) on the philosophy of science with a sample of case studies from various fields."

Nageswari Rajaratnam ("A Model for the Comparison of Individual and Group Performance Scores"), a native of Ceylon, is an assistant professor in the College of Education at the University of British Columbia. Educated at the University of London, he came to this country and received his Ph.D. degree from the University of Illinois in 1957. He writes that he is presently engaged in research on the theory of measurement in collaboration with Dr. Lee J. Cronbach and Dr. Goldine Gleser, and is also working on a study of the effects of various factors on transfer and retention. Dr. Rajaratnam relates that other interests center on developing better and more useful techniques for doing research and interpreting results in the social sciences, and on investigations related to concept formation and problem-solving behavior of human subjects. He has recently published articles in the field of measurement theory.

Beatrice K. Rome ("The Leviathan Technique for Large-Group Analysis") is a member of the planning staff in the System Training Operations Department at the System Development Corporation, Santa Monica, California. A native of the Russian Ukraine, she received her formal education at Radcliffe and Harvard where she earned her B.A. and Ph.D. degrees in philosophy, carrying on further study at the Sorbonne and College de France. In addition to her research on the Leviathan studies, she writes that she is also interested in problems of conflict resolution and in Soviet strategy in particular. A recent article on "NATO's Dilemna: The Soviet Attitude Toward and Response to NATO, Soviet Conflict Techniques" reflects her work in this interest area. The background and origin of the Leviathan studies, and the work being carried on in this research, is discussed below in connection with Dr. Sydney Rome.

Sydney C. Rome ("The Leviathan Technique for Large-Group Analysis"), relating how he and his wife originated the Leviathan studies, writes, "The Leviathan studies were born, appropriately enough, on account of the family plan. We had been invited to talk on "intentionally structured systems" in New York and elected to go by train, necessitating a Sunday departure. Out west, family plan begins on Mondays and ends Thursdays. Therefore, we left the preceding Thursday and went via Portland, Oregon. If the logic of our choice is obscure, note that we do like to be together on good trains. By the time we had reached New York, our ideas were fairly clear. Our practical problem was to develop a program of systems research that would go back to Plato and Parmenides for a philosophy of organism and that would include pertinent reference to general systems theory on the parts of Aristotle and Aquinas, Descartes and Spinoza, and Bradley, Bergson, and Whitehead. What all these have in common, for us, is a major concern for the problem of the one and the many. What concepts and theoretical syntheses are adequate for understanding the interaction of well-developed hierarchical or morphological structure with dynamic change? In addition, we wanted our research to benefit from our experience at the RAND and System Development Corporations both in simulating the significant qualitative aspects of the total air defense nexus on digital computers and also in incorporating computers in real time in the command and control of large social organizations. It took a rather large concept to unify this much background, but students of political philosophy may grant that this is not the overt reason why we call the project Leviathan. Still, as one wag has put it, everything about Leviathan is titanic (and you know what happened to the Titanic). In any case, the demands of the project have led us into recondite areas, such as a series of basic definitions invariant to varieties of pos-

sible sociological theories, coding social interactions into computer-oriented language, the development of analogue techniques for simulating social process on digital computing machinery, and the application of informational multivariate analysis to formal social structure. While the Mark I programs have been checked out as a system and are operating, thus proving that the Leviathan-type analogue simulation is feasible on a digital machine, they are almost inaccessible. What has not happened yet, although it is momentarily due, is to be able to read out in intelligible quantitative terms what in the world happens when we run our complex programs on the computer." Dr. Rome received his education at Harvard, gaining his B.S. in mathematics and his Ph.D. in philosophy. At present on the Technical Staff, Research Directorate, at System Development Corporation, he and Dr. Beatrice Rome have published several articles on intentionally structured systems and their Leviathan studies.

Ivan D. Steiner ("A Model for the Comparison of Individual and Group Performance Scores"), associate professor and associate head of the department of psychology at the University of Illinois, received his M.A. and Ph.D. degrees from The University of Michigan. He has since held a postdoctoral fellowship in personality at Illinois and a Ford Faculty Fellowship for research in group processes and productivity. The author of numerous articles, Professor Steiner relates that he is "continuing work on models with which to 'predict' the productivity of groups, and am doing research into dissonance, incongruity, or asymmetrical A-B-X situations (depending upon whose language is preferred)." He adds that with regard to the present paper "the problem became central in my thinking when I spent a year (1959–60) reading, and attempting to integrate, some of the research dealing with group processes and productivity. It became apparent that the experimental tasks which have been employed by different researchers have often permitted or required very different patterns of resource usage, and that this difference is often neglected in discussions of group productivity. Especially in the case of research comparing the performance of individuals and groups, a failure to take account of the ways in which the resources of group members can be combined had led to many seemingly contradictory findings. It became apparent also that in many studies it is more pertinent to ask whether groups perform better than their most competent members would have performed if they had worked alone than to ask whether groups perform better than the average individual who works alone. In discussions with Nageswari Rajaratnam, then a research associate with Dr. Lee Cronbach, it became clear that the procedures recommended in this paper might be used to good advantage."

Anders Sweetland ("An Analysis of the Decision-Making Functions of a Simulated Air Defense Direction Center") is a psychologist at the RAND Corporation, Santa Monica, California. The recipient of a Ph.D. degree from the University of North Dakota, Dr. Sweetland is currently interested in systems research, and in manned simulations in particular. He is the author of numerous articles, which have appeared primarily in psychological journals, on scaling methods and problems, the most recent one being "Traits of fatherhood as revealed by the factor analysis of a parent attitude scale."

Clyde M. Williams ("Models for Medical Diagnosis") received his Ph.D. from Oxford University in 1954 following medical training at Baylor University College of Medicine where he earned an M.D. in 1952. Now a special trainee in radiology at the University of Florida College of Medicine, his present research interests are in the use of multivariate techniques for the diagnosis of thyroid disease and in the intermediary metabolism of catecholamines. The author of 35 papers and a book on *Thyroid Scanning*, Dr. Williams tells of his work with Dr. Overall (co-author of the present article) during the past years on the application of multivariate statistical techniques to biological problems and, in particular, to medical diagnosis. He goes on to say, "At the present time, we are collaborating on the use of statistical techniques for the diagnosis of thyroid disease, making use of a 20% sample of all cases of thyroid disease diagnosed in the Veterans Administration in 1958. The primary object of this problem is to determine whether statistical classification procedures can be utilized to effect greater precision in this area of medical diagnosis."

Herbert Zimmer ("Preparing Psychophysiologic Analog Information for the Digital Computer") received his A.B. degree from Western Reserve University in 1948 and his Ph.D. in clinical psychology from the University of Rochester in 1952, where he was a V.A. trainee. At present an associate professor of psychology at the University of Georgia, Dr. Zimmer previously worked as a research psychologist for the Air Research and Development Command and as an assistant professor at Georgetown University. His interest in computers stems from his realization that research in psychophysiology presents "a big data reduction and analysis problem." The system described in his paper was developed at the District of Columbia General Hospital with technical guidance from the National Bureau of Standards and under the sponsorship of the Air Force Office of Scientific Research.

WILEY BOOKS

SYSTEMS: RESEARCH AND DESIGN

Proceedings of The First Systems Symposium at the Case Institute of Technology

Edited by DONALD P. ECKMAN, *Case Institute of Technology*. This book has been prepared by some of the most prominent figures in the systems engineering field. It contains fourteen papers, most of which deal with analysis and synthesis in very large complex systems such as industrial manufacturing complexes, military systems, economic systems, and systems involving both men and machines. Among the basic topics discussed in these papers are systems theory, communication, control, performance criteria, optimization, reliability, and human functions in the systems. *1961. Approx. 328 pages. Prob. $9.00.*

PROGRESS IN OPERATIONS RESEARCH

Volume I

Edited by RUSSELL L. ACKOFF, *Case Institute of Technology*. The inaugural volume of a series of review volumes which will inventory mathematical techniques and research methods available to operations research. The series is designed to serve as a basic source for reliable, up-to-date information on major OR techniques. The primary emphasis of Volume I is on technical progress (i.e., the development of modelling techniques and their application to problem-solving) in eight important areas of OR: Inventory Theory, Linear Programming, Dynamic Programming, Queuing Theory, Sequencing Theory, Replacement Theory, Simulation, and Gaming. *1961. Approx. 518 pages. Prob. $9.75.*

ORGANIZED EXECUTIVE ACTION

Decision-Making, Communication, and Leadership

By HENRY H. ALBERS, *University of Iowa*. In this book the author examines the basic elements of executive action within the hierarchical framework of a business organization and provides insights into the way in which diverse skills and techniques become integrated into the managerial process. The book draws information from many different fields, e.g., economics, group dynamics, operations research, information theory, semantics, accounting, quality control, and electronic data processing. *1961. 586 pages. $8.50.**

PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE ON OPERATIONS RESEARCH

Edited by JOHN BANBURY *and* JOHN MAITLAND. Containing the proceedings of a conference held in 1960 at Aix-en-Provence, France, this volume offers a unique view of progress in operations research throughout the world. It contains the contributions of experts from India, the Soviet Union, Japan, Israel, the Netherlands, Belgium, France, the United Kingdom, Canada, and the United States. *1961. In press.*

* *Also available in a textbook edition.*

Send for examination copies.

JOHN WILEY & SONS, Inc.

440 Park Avenue South, New York 16, N.Y.